BS

12

AFGL-TR-81-0028

LEVEL II

ADA104036

MILITARY GEODESY AND GEOSPACE SCIENCE
Unit One

Warren G. Heller
A. Richard LeSchack

The Analytic Sciences Corporation
One Jacob Way
Reading, Massachusetts 01867

February 1981

Scientific Report No. 5

Approved for public release; distribution unlimited

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

81 9 10 038

BLANK PAGES IN THIS DOCUMENT WERE NOT FILMED

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFGL-TR-81-0028	2. GOVT ACCESSION NO. AD-A104 038	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MILITARY GEODESY AND GEOSPACE SCIENCE Unit One		5. TYPE OF REPORT & PERIOD COVERED Scientific [redacted] 5.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Warren G. Heller A. Richard LeSchack		8. CONTRACT OR GRANT NUMBER(s) F19628-77-C-0152
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Analytic Sciences Corporation One Jacob Way Reading, Massachusetts 01867		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 320432AA
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Monitor/Brian Mertz, Capt, USAF/LWG		12. REPORT DATE Feb[redacted]981
		13. NUMBER OF PAGES 194
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Geodesy, gravity, physical geodesy, mapping, charting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This lecture course provides a full-year introduction to Military Geodesy and Geospace Science. Throughout the presentation a military perspective is maintained which links Mapping, Charting, and Geodesy (MC&G) issues with modern defense requirements. Elementary preparation is assumed in the subjects of general physics, mechanics, chemistry, astronautics, and linear system theory. The student should also be familiar with differential equations, analytic geometry, and linear algebra. Some acquaintance with vector calculus is useful but not essential.

The topics covered herein are intended to provide conceptual rather than working knowledge. Ideally, the student completing this course will have attained a broad understanding of the MC&G field and will be able to develop specialized expertise quickly when required.

The notes are intended to be presented in chapter/section order within each of the four Units of Instruction. However, several of the subsections in these notes contain more advanced material which may be omitted without loss of continuity. These subsections are denoted with the symbol (†) after the title. A fifth volume contains faculty material.

The organizational flow of the lectures is from concepts in the initial sections, particularly in Unit One, to applications and specific systems later on. As a result the student is often referred ahead to provide motivation in regard to relevancy. In later chapters, however, the situation is reversed with the student referred back to review important conceptual material as necessary.

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A	

FOREWORD

This lecture course provides a full-year introduction to Military Geodesy and Geospace Science. Throughout the presentation a military perspective is maintained which links Mapping, Charting, and Geodesy (MC&G) issues with modern defense requirements. Elementary preparation is assumed in the subjects of general physics, mechanics, chemistry, astronautics, and linear system theory. The student should also be familiar with differential equations, analytic geometry, and linear algebra. Some acquaintance with vector calculus is useful but not essential.

The topics covered herein are intended to provide conceptual rather than working knowledge. Ideally, the student completing this course will have attained a broad understanding of the MC&G field and will be able to develop specialized expertise quickly when required.

The notes are intended to be presented in chapter/section order within each of the four Units of Instruction. However, several of the subsections in these notes contain more advanced material which may be omitted without loss of continuity. These subsections are denoted with the symbol (†) after the title.

The organizational flow of the lectures is from concepts in the initial sections, particularly in Unit One, to applications and specific systems later on. As a result the student is often referred ahead to provide motivation in regard to relevancy. In later chapters, however, the situation is reversed with the student referred back to review important conceptual material as necessary.

GLOSSARY OF ACRONYMS AND ABBREVIATIONS FOR UNIT ONE

AENA	-	American Ephemeris and Nautical Almanac	1-58
ANNA-1B	-	Satellite launched in 1962 (the acronym represents Army, Navy, NASA, Air Force)	1-140
ATS-6	-	Applications Technology Satellite Number Six (a radio relay satellite)	1-153
BGM-2	-	Bell Gravity Meter (Model two)	1-136
BIH	-	Bureau International de l'Heure	1-53
CIO	-	Conventional International Origin (north pole)	1-49
DMA	-	Defense Mapping Agency	iii
DoD	-	Department of Defense	1-152
DMAHTC	-	Defense Mapping Agency Hydrographic Topographic Center	1-82
DUT1	-	Difference between UT1 and UTC time scales	1-50
ED	-	European Datum	1-73
FGCC	-	Federal Geodetic Control Committee	1-13
GEOCEIVER	-	Geodetic (radio position determining) receiver	1-151
GEOS	-	Geodetic Earth Orbiting Satellite	1-140
GEM 10B	-	Goddard Earth Model	1-161
GRS 67	-	Geodetic Reference System, 1967	1-6
HIRAN	-	High Precision SHORAN	1-16
IF	-	Intermediate Frequency	1-113
ILS	-	International Latitude Stations	1-54
INS	-	Inertial Navigation System	1-104
IPMS	-	International Polar Motion Service	1-53

GLOSSARY OF ACRONYMS AND ABBREVIATIONS FOR UNIT ONE (Continued)

LAGEOS	-	Laser Geodynamics Satellite	1-140
LO	-	Local oscillator	1-113
MC&G	-	Mapping, Charting, and Geodesy	1-1
MMD 68	-	Modified Mercury Datum 1968	1-73
NAD 27	-	North American Datum 1927	1-73
NASA	-	National Aeronautics and Space Administration	1-152
NSWC	-	Naval Surface Weapons Center	1-160
Pulkovo 42	-	Russian Geodetic Datum	1-73
PZT	-	Photographic Zenith Tube	1-52
RF	-	Radio frequency	1-113
RMS	-	Root Mean Square	1-84
SAO	-	Smithsonian Astrophysical Observatory	1-152
SAO 66-C6	-	Smithsonian Datum 1966	1-73
SEASAT	-	Sea Satellite	1-140
SECOR	-	Sequential Collation of Range	1-148
SHIRAN	-	S-Band High Accuracy Ranging and Navigation (a radio distance measuring system)	1-16
SHORAN	-	Short Range Navigation	1-16
TAI	-	International Atomic Time	1-51
TD	-	Tokyo Datum	1-73
UT	-	Universal Time	1-44
UTC	-	Coordinated Universal Time	1-50
UTM	-	Universal Transverse Mercator (a map projection)	1-96

GLOSSARY OF ACRONYMS AND ABBREVIATIONS FOR UNIT ONE (Continued)

UT0	-	Universal Time determined directly from star observations	1-49
UT1	-	Universal Time corrected to an internationally agreed upon mean polar position	1-49
UT2	-	Universal Time corrected for the seasonal variation in rotational speed, determined empirically from past observations	1-49
VLBI	-	Very-long baseline interferometry	1-113
VZT	-	The visual zenith telescope	1-52
WGS 72	-	World Geodetic System 1972	1-6
ZUPT	-	Zero velocity update	1-105

UNIT ONE

TABLE OF CONTENTS

UNIT ONE

TABLE OF CONTENTS (Continued)

UNIT ONE

TABLE OF CONTENTS (Continued)

UNIT ONE

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

UNIT ONE

LIST OF TABLES

UNIT ONE

INTRODUCTION TO MAPPING, CHARTING, AND GEODESY (MC&G)

CHAPTER ONE
INTRODUCTION

Unit One introduces the student to Mapping, Charting, and Geodesy. Basic concepts and principles are presented that will be applied during the remainder of the course. The subjects to be covered include:

- Earth modeling
- Coordinate systems
- Techniques of mapping, charting, and geodesy.

All of these areas are part of the science of geodesy, which is defined by the three principal subjects with which it is concerned:

- The size and shape of the earth
- The relative location of points on or near the surface of the earth
- The earth's gravity field.

The first two areas are referred to as geometric geodesy; the third is physical geodesy. The geometric and physical (or gravitational) aspects of geodesy are closely related to one

another, since the physical surface of the earth does not deviate greatly from an equipotential* surface of the gravity field.

The material of Unit One is organized into three chapters, each examining geodesy from a different point of view:

- Geometric geodesy (Chapter Two) -- including material on relevant aspects of cartography and surveying.
- Physical geodesy (Chapter Three)
- Satellite geodesy (Chapter Four) -- emphasizing the unique contributions to geodesy resulting from the use of earth satellites.

*The physical concepts of field and potential are reviewed in Appendix A.

CHAPTER TWO
GEOMETRIC GEODESY

1.2.1 The Ellipsoid and the Geoid

The Ellipsoid

For many purposes, it is adequate to consider the earth to be a sphere. When a more precise model is required, a mathematically defined surface called a spheroid is used to model the deviation of the earth from true spherical form, consisting principally of a flattening at the poles and bulging of the equator by about one part in 300. Many sphere-like surfaces have been used, at one point or another, some of them extremely complicated in their mathematical form. Since about 1930 it has been standard practice to use a relatively simple surface, the ellipsoid of revolution (Fig. 1.2-1), to model the earth. Considerable effort has been expended to deduce values for the shape and size of the ellipsoid that best describes the earth. Table 1.2-1 lists some of the ellipsoids that have been used at various times.

It is important to note that the ellipsoid is not only a model for the shape of the earth, but for the gravity field as well, since the ellipsoid can be considered a surface of equal potential* for which a theoretical or normal value of gravity can be determined. Normal gravity provides a general approximation to the earth's actual gravity field. Gravity field modeling will be discussed further in Section 1.3.2

Points on the surface of an ellipsoid can be located in terms of their ellipsoidal coordinates, as shown in Fig. 1.2-2. The geodetic latitude ϕ, at point P, is defined as

*The concept of potential is reviewed in Appendix A.

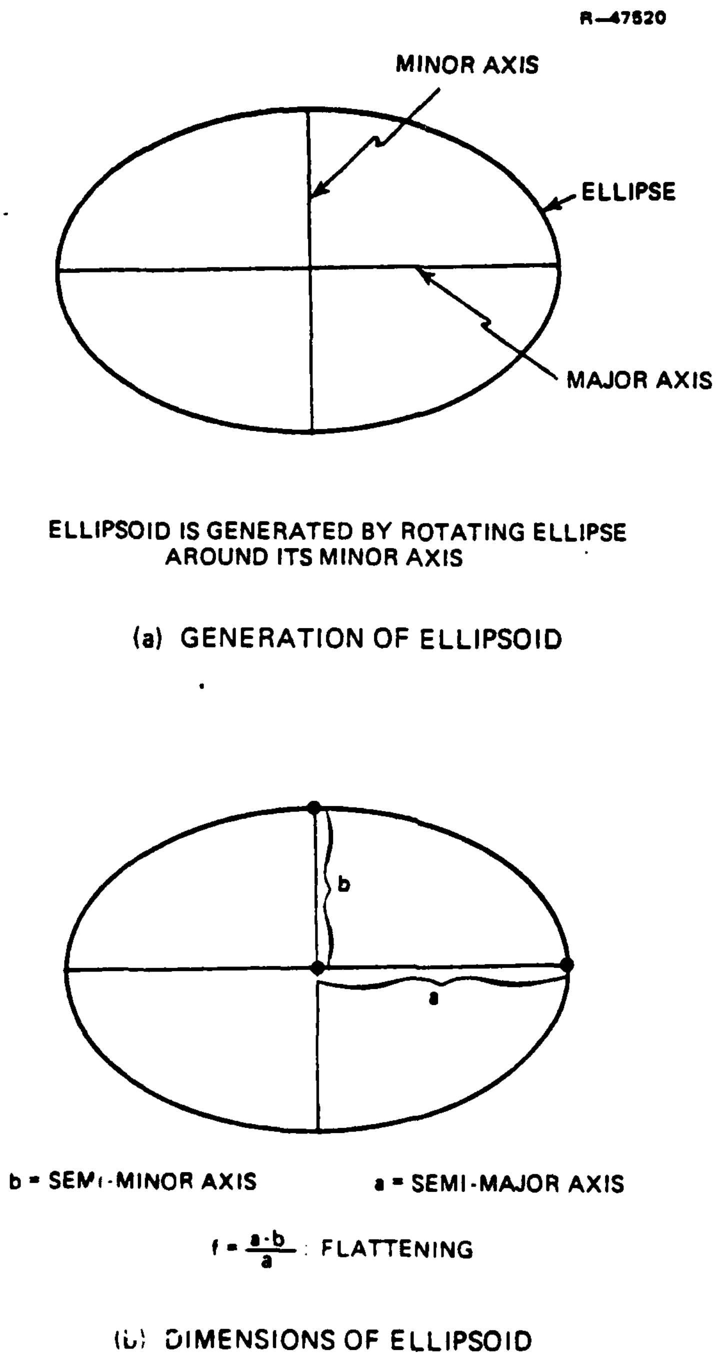
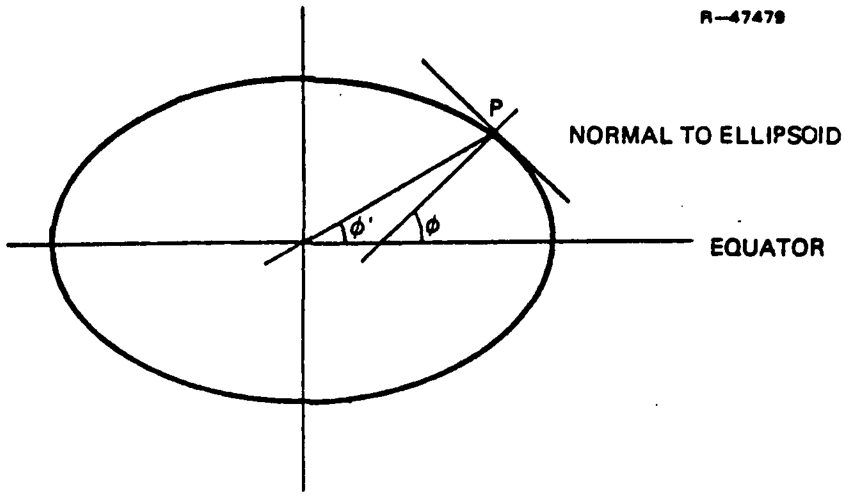
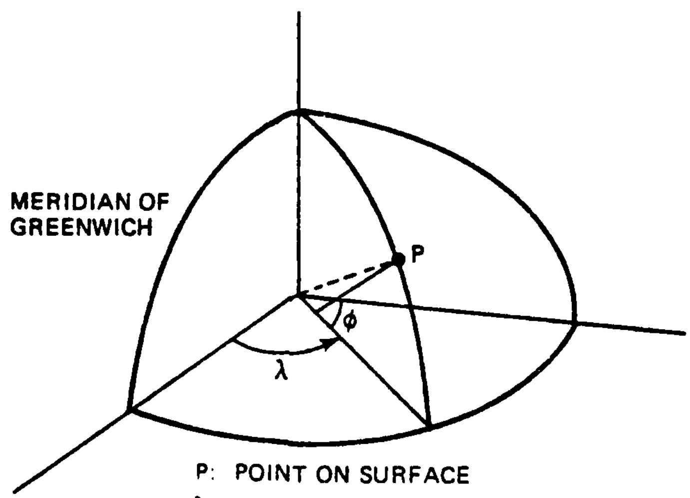


Figure 1.2-1 The Ellipsoid of Revolution



P: POINT ON SURFACE
φ: GEODETIC LATITUDE
φ': GEOCENTRIC LATITUDE

(a) LATITUDE



P: POINT ON SURFACE
λ: GEODETIC LONGITUDE

(b) LONGITUDE

Figure 1.2-2   Ellipsoidal (Geodetic) Coordinates

TABLE 1.2-1

EARTH ELLIPSOIDS

DESIGNATION	SEMI-MAJOR AXIS (m)	FLATTENING
Airy (1830)	6377563	1/299.32
Bessel (1841)	6377397	1/299.15
Clarke (1866)	6378206	1/294.98
Clarke (1880)	6378249	1/293.46
International (1924)	6378388	1/297.00
Kaula (1961)	6378165	1/298.30
Smithsonian (1966)	6378165	1/298.25
Geodetic Reference System (GRS 67)	6378160	1/298.25
World Geodetic System (WGS 72)	6378135	1/298.26

the angle between the normal to the ellipsoid at point P and the equatorial plane. The geocentric latitude ϕ', used in some applications, is defined as the angle at the center of the ellipsoid between the plane of the equator and a line to point P. The mathematical relationship between ϕ and ϕ' is given by the formula:

$$\tan \phi' = (1-f)^2 \tan \phi \qquad (1.2\text{-}1)$$

Sometimes it is more convenient to have a direct expression for the difference $(\phi-\phi')$:

$$\phi-\phi' = (f+\frac{1}{2}f^2) \sin 2\phi - \frac{1}{2}f^2 \sin 4\phi \qquad (1.2\text{-}2)$$

The maximum value of $\phi-\phi'$ is about 1/5 deg.

The Geoid

Another basic reference surface is the geoid,* a level (or equipotential) surface of the earth's gravity field. The ocean surface (mean sea level) corresponds to the geoid. The direction of the gravity vector at any point (plumb line or vertical) is normal to the geoid. The relation between the geoid and ellipsoid surfaces is shown schematically in Fig. 1.2-3. Of particular interest are two features of the diagram:

- The difference in direction between the normal to the geoid (which is the direction of the gravity vector) and the normal to the ellipsoid. This difference in direction is known as the deflection of the vertical.
- The vertical separation between the geoid and ellipsoid surfaces, known as the geoid height or undulation.

These concepts (to be studied in more detail in Chapter Three) are used frequently throughout the text.

1.2.2 Geodetic Positioning

Geodetic positioning refers to the determination of the geodetic coordinates (latitude, longitude, and height) of various points on or near the earth's surface, for the purpose of defining precisely the spatial relationships that exist among those points. Reduced to its simplest terms, the goal is this: Given points A and B, answer the questions:

- How far?
- What direction?

---

*To be studied in more detail in Section 1.3.1.

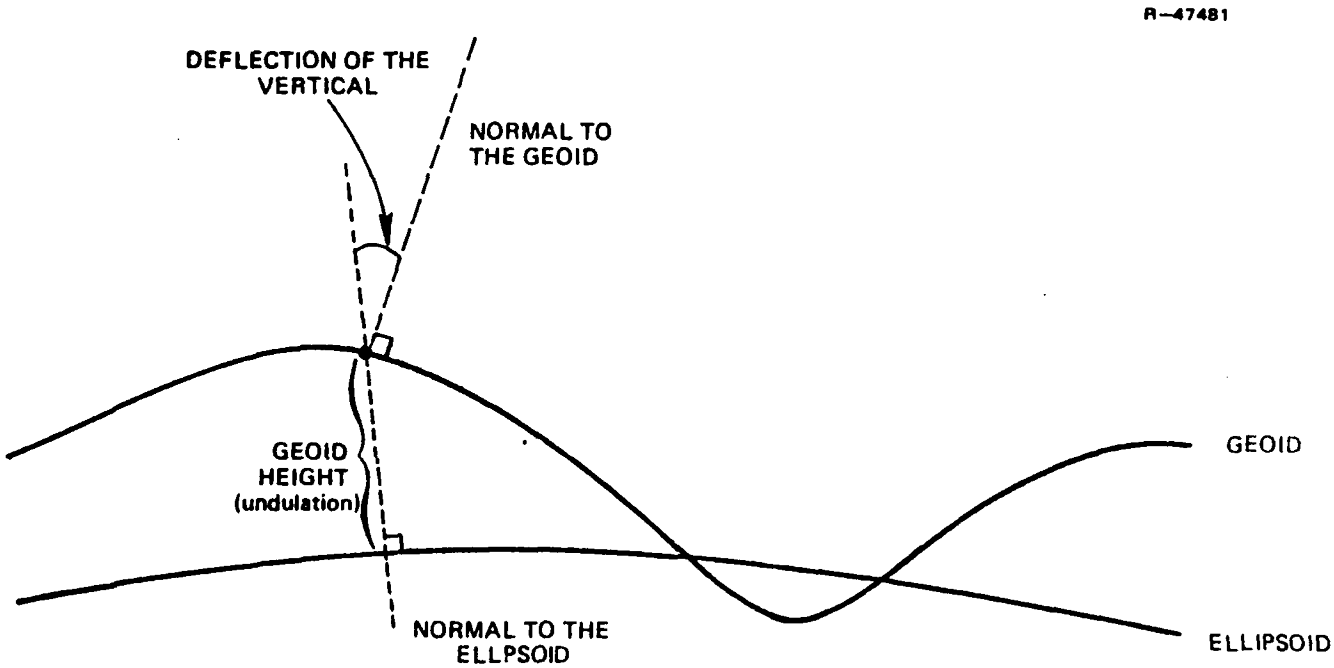


Figure 1.2-3 Relation Between Geoid and Ellipsoid

Although perhaps not usually thought of in exactly these terms, maps and navigational charts of all kinds are, in effect, analog devices for answering such questions. The geodetic coordinates are expressed with respect to a particular ellipsoid, as shown in Fig. 1.2-4 and 1.2-5, and are measured in a particular coordinate system (called a datum), based on an adopted origin (for which the geodetic coordinates are known or assumed). The subject of datums will be covered in detail in Section 1.2.5.

Datums were originally developed on a purely local basis, and then extended to provide uniform coverage of an entire region, like a continent. However, it was not possible, in the pre-satellite era, to provide accurate ties between one datum and another, or to develop a unified worldwide geodetic system in which the coordinates of widely separated points could be expressed with great precision. Intensive efforts have been directed, since the end of the 1950s, toward these ends; some of the details are discussed in Unit 2, Section 2.2, under the heading World Geodetic Systems.

The basic techniques of geodetic positioning are divided into:

- Horizontal control, referring to the determination of latitude and longitude
- Vertical control, referring to the determination of height

and different approaches have been developed in these two areas. Three methods of horizontal control, triangulation, trilateration, and traverse, are discussed in the next sections. Treatment of vertical control methods follows.

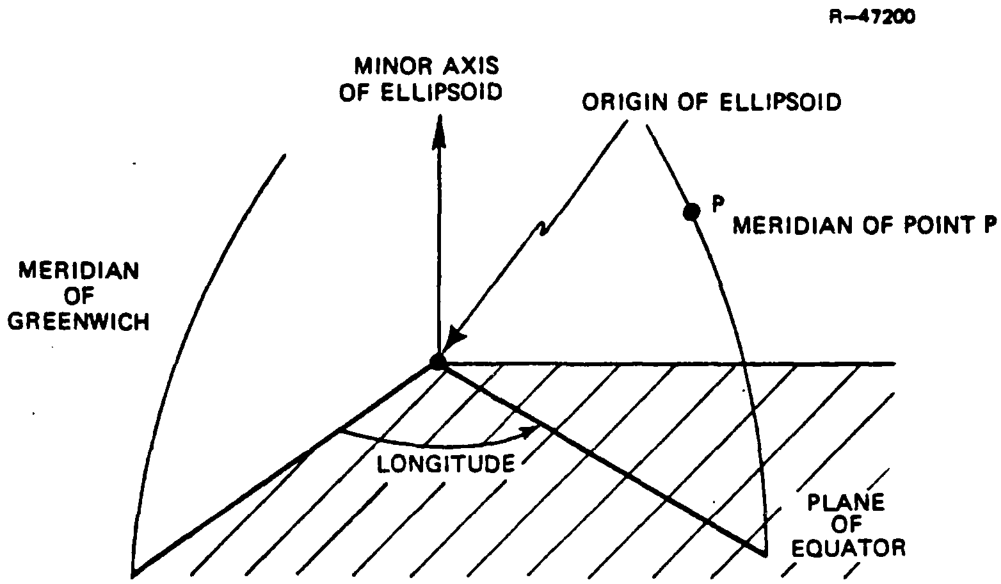


Figure 1.2-4 Geodetic Longitude

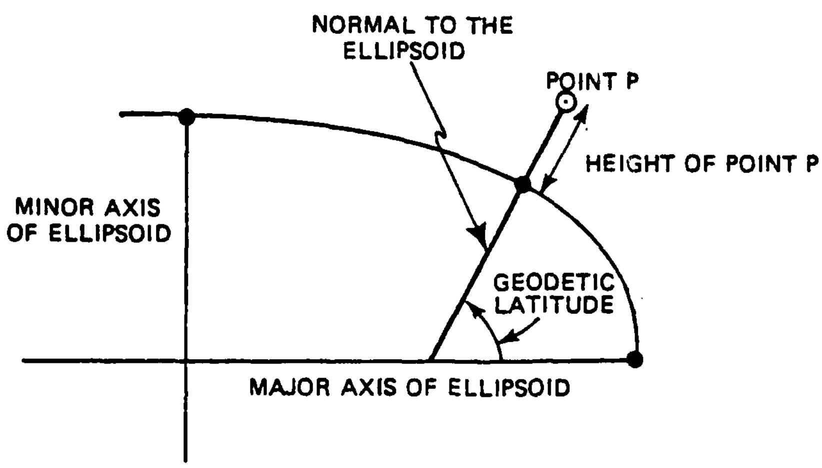


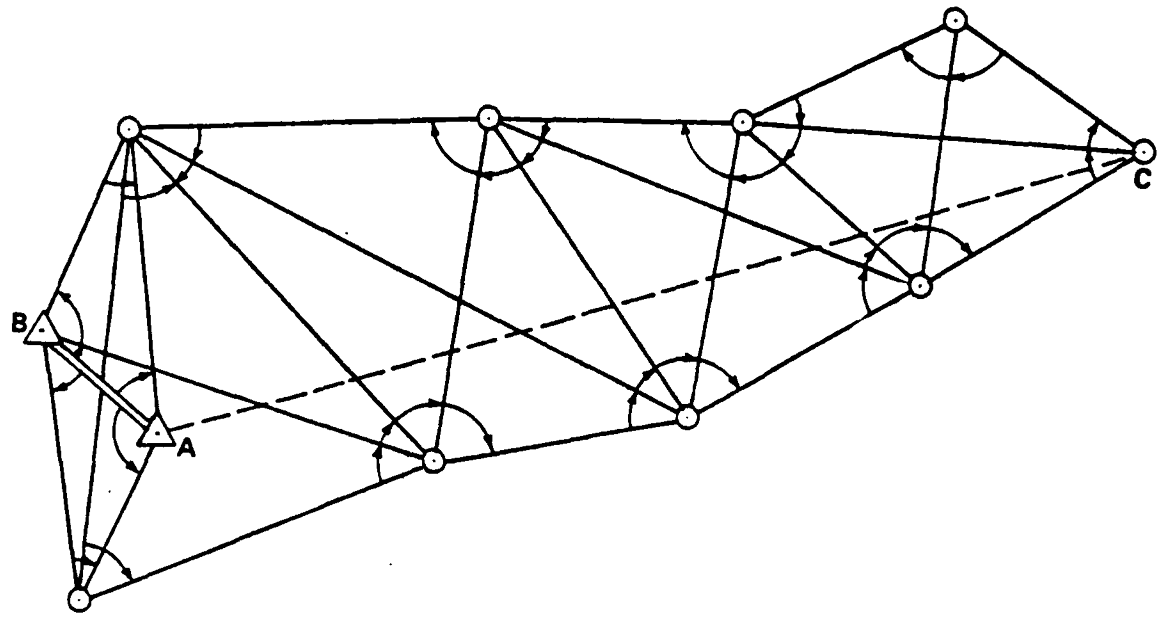
Figure 1.2-5 Geodetic Latitude and Height

Triangulation - Triangulation, the most common form of geodetic survey (for horizontal control), consists of the measurement of the angles of a series of triangles. The principle of triangulation is based on simple trigonometric procedures. If the length of one side of a triangle and the angles at each end of the side are accurately measured, the other two sides and the remaining angle can be computed. Normally, all of the angles of every triangle are measured to furnish exact data for use in computing the precision of the measurements (Fig. 1.2-6). Also the known geodetic latitude and longitude at one end of the measured side, along with the length and direction (azimuth) of the side, provide sufficient data to compute the latitude and longitude of the other end of the measured side.

The measured side of the basic triangle is called a base line. Measurements are made as carefully and accurately as possible with specially calibrated tapes or wires of invar, an alloy highly resistant to changes in length resulting from changes in temperature. The tapes or wires are checked periodically against standard measures of length (at the Bureau of Standards in the United States and corresponding agencies in other countries). The geodimeter and tellurometer, operating on electro-optical and electronic principles respectively, have replaced the older methods of base measurement in the more precise surveys. Using such equipment the work of measuring distances can be completed more rapidly and accurately than with wire or tape. The laser-equipped geodimeter has proven to be the most accurate, particularly over long distances.

To establish an arc of triangulation between two widely separated locations, a base line may be measured and longitude and latitude determined for the initial points at one

R—47202



KNOWN DATA:
Length of baseline AB.
Geodetic latitude and longitude of points A and B.
Azimuth of line AB.

MEASURED DATA:
Angles to new control points.

COMPUTED DATA:
Geodetic latitude and longitude of point C, and other new points.
Length and azimuth of line AC.
Length and azimuth of all other lines.

Figure 1.2-6 Example of a Simple Triangulation Net

end. The locations are then connected by a series of adjoining triangles forming quadrilaterals (four-sided figures) extending from each end, as shown in Fig. 1.2-6. All angles of the triangles are observed repeatedly to reduce errors. With the longitude, latitude, and azimuth of the initial points, similar data can be computed for each vertex of the triangles. This establishes, at each of these points, a <u>triangulation</u> or

geodetic control station. The coordinates of each of the stations are defined as geodetic coordinates.

Triangulation is extended over large areas by connecting and extending series of arcs and forming a network or triangulation system. The network is adjusted in a manner which reduces the effect of observational errors to a minimum. A denser distribution of geodetic control is achieved in a system by subdividing or filling in with other surveys. Major triangulation networks have been established over large parts of the land surface of the earth.

There are four levels, or categories, of triangulation, based on accuracy requirements. Uniform specifications for these levels of triangulation are established as part of a broader set of standards and procedures known as the "Classification, Standards of Accuracy, and General Specifications of Geodetic Control Survey," developed by an interagency Federal Geodetic Control Committee (FGCC). First-Order (Primary Horizontal Control) is the most accurate triangulation. It is costly and time-consuming, because it uses the best instruments and the most rigorous computation methods. First-Order triangulation is used to provide the basic framework of horizontal control for a large area (such as for a national network). It has also been used in preparation for metropolitan expansion and for scientific studies requiring exact geodetic data. Its accuracy should be at least one part in 10^5.

Second-Order, Class I (Secondary Horizontal Control) includes the area networks between the First-Order arcs and detailed surveys in very high-value land areas. Surveys of this class strengthen the U.S. National Horizontal Control Network and are adjusted as part of the network. Therefore, this class also includes the basic framework for further

densification. Second-Order, Class I triangulation should maintain an accuracy of at least one part in 5×10^4.

The demands for reliable horizontal control surveys in areas that are not highly developed (and where no development is expected in the near future) create a need for triangulation classified as Second-Order, Class II (Supplemental Horizontal Control). This class is used to establish control along the coastline, inland waterways, and interstate highways. The control data contribute to the National Network and are published as part of the network. The minimum accuracy allowable in Class II of Second-Order is one part in 2×10^4.

Third-Order, Class I and Class II (Local Horizontal Control), is used to establish control for local improvements and developments, topographic and hydrographic surveys, or for other projects requiring moderate accuracy. This triangulation is carefully connected to the National Network. The work should be performed with sufficient accuracy to satisfy the standards of one part in 10^4 for Class I and one part in 5×10^3 for Class II. Spires, stacks, standpipes, flag poles, and other identifiable objects located to this degree of accuracy also have significant value for many surveying and engineering projects.

The sole accuracy requirement for Fourth-Order triangulation is that the positions be located without any appreciable error on maps compiled on the basis of the control.

The accuracy standards for the various orders of triangulation are summarized in Table 1.2-2.

TABLE 1.2-2

ACCURACY STANDARDS FOR ORDERS OF TRIANGULATION*

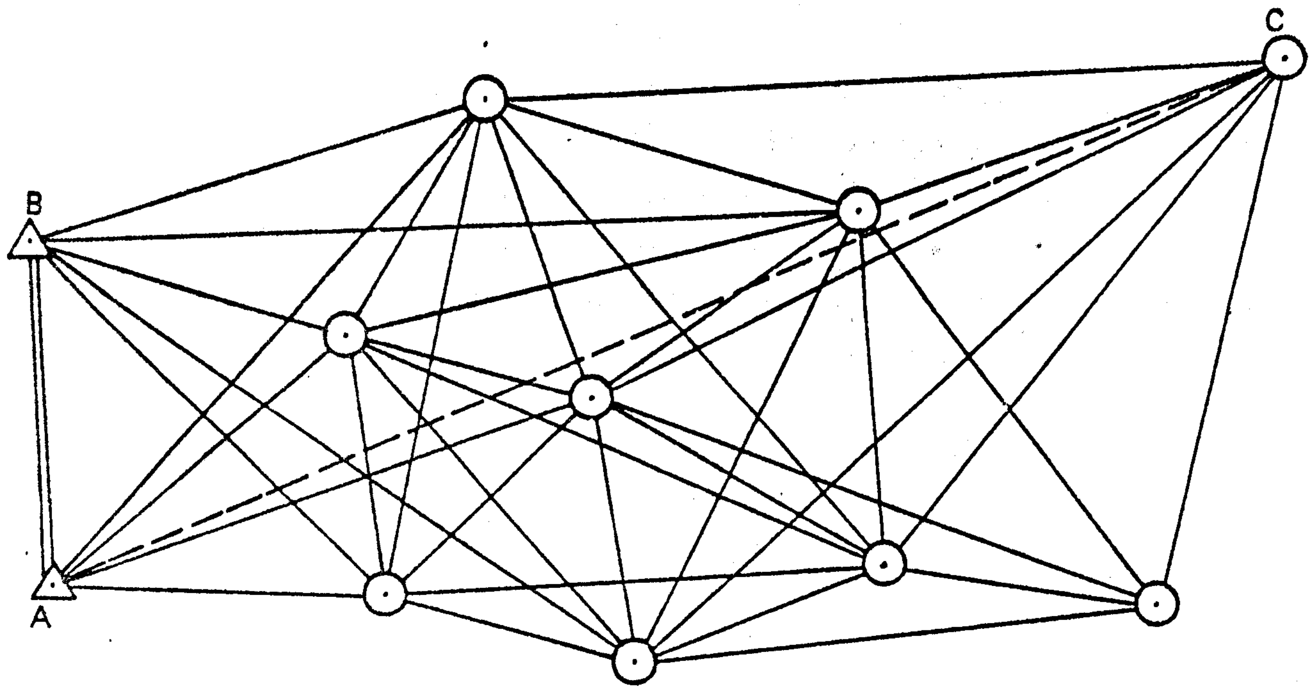
ORDER	ACCURACY REQUIREMENT
First Order	One part in 10^5 or better
Second Order	
Class I	One part in 5×10^4
Class II	One part in 2×10^4
Third Order	
Class I	One part in 10^4
Class II	One part in 5×10^3
Fourth Order	Sufficient for purposes of map construction

Normally, triangulation is carried out by parties of surveyors occupying preplanned locations (stations) along the arc and making all the measurements as they proceed. When distances between two points are too long for conventional methods, connections were sometimes made, in the past, by a method known as flare triangulation. Stations are occupied on either side of the gap and flares or beacons are parachuted from aircraft or shot into the air from ships at suitable points between the stations. Intersections of lines are made simultaneously at all of the stations and reasonably accurate bridges established. Historically, a pioneering connection of this type was established between Norway and Denmark. Flare triangulation is now obsolete, and much longer gaps are now bridged routinely by using modern techniques such as Satellite Geodesy (Chapter Four).

*Extracted from publications of the Federal Geodetic Control Committee.

Trilateration - The concept of trilateration is essentially the same as triangulation, except that it is based exclusively on distance measurements and involves no measurement of angles. Figure 1.2-7 illustrates the basic technique. In older surveying practice, triangulation was more important than trilateration because there were no practical ways of measuring great distances directly, to the required degree of accuracy, while precise measurement of angles was feasible. Modern methods of electronic distance measuring make trilateration a useful alternative over distances on a continental scale. Two examples of modern distance measuring technology -- the geodimeter and the tellurometer -- have already been mentioned in the discussion of triangulation techniques. The geodimeter, a laser device, has been used, for example, to attain internal accuracies better than one part per million for transcontinental surveys in the United States. These surveys consist of a series of high-precision length, angle, and astronomic azimuth determinations running approximately east-west and north-south through the conterminous states. The tellurometer is a microwave device; an example of its use is the completion of the Australian Geodetic Datum network covering that continent. With an average loop length of about 1500 km, the internal accuracy was of the order of two to three parts per million.

Other examples of electronic distance measuring systems are SHORAN, HIRAN, and SHIRAN, also used for purposes of aircraft navigation and reconnaissance. These systems measure long lines (up to 800 km), permitting the extension of geodetic triangulation networks over vast areas in comparatively short periods of time. In addition, the surveys of islands and even continents separated by extensive water barriers have been connected by trilateration based on these techniques. The Canadian SHORAN Network, connecting the sparsely-populated



KNOWN DATA:
Length of baseline AB.
Latitude and longitude of points A and B.
Azimuth of line AB.

MEASURED DATA:
Length of all triangle sides.

COMPUTED DATA:
Latitude and longitude of point C, and other new points.
Length and azimuth of line AC.
Length and azimuth between any two points.

Figure 1.2-7 Example of a Trilateration Net

northern coastal and island areas with the central part of the country, and the North Atlantic HIRAN Network, tying North America to Europe, are further examples of trilateration. Many other trilateration networks (SHORAN and HIRAN) have been established throughout the world. SHIRAN has been used in the interior of Brazil.

Only distances are measured in trilateration and each side is measured repeatedly to insure precision. The entire network is then adjusted to minimize discrepancies. The angles of the triangles are computed, and geodetic positions are obtained, as in triangulation, for the stations at the vertices of the triangles.

When baseline distances are large, it is essential to make precise corrections for a number of effects to preserve the inherent accuracy of the electronic or laser devices. These include:

- Correction for effective velocity of propagation of electromagnetic energy through the atmosphere, as influenced by pressure, temperature, and relative humidity
- Correction for path length changes because of refraction, caused by changes in the index of refraction (propagation velocity) from one place to another
- Geometric corrections to relate the ray path distance (straight line, except as affected by refraction) to a corresponding arc length on the ellipsoid.

Refraction and geometric corrections are shown schematically in Fig. 1.2-8.

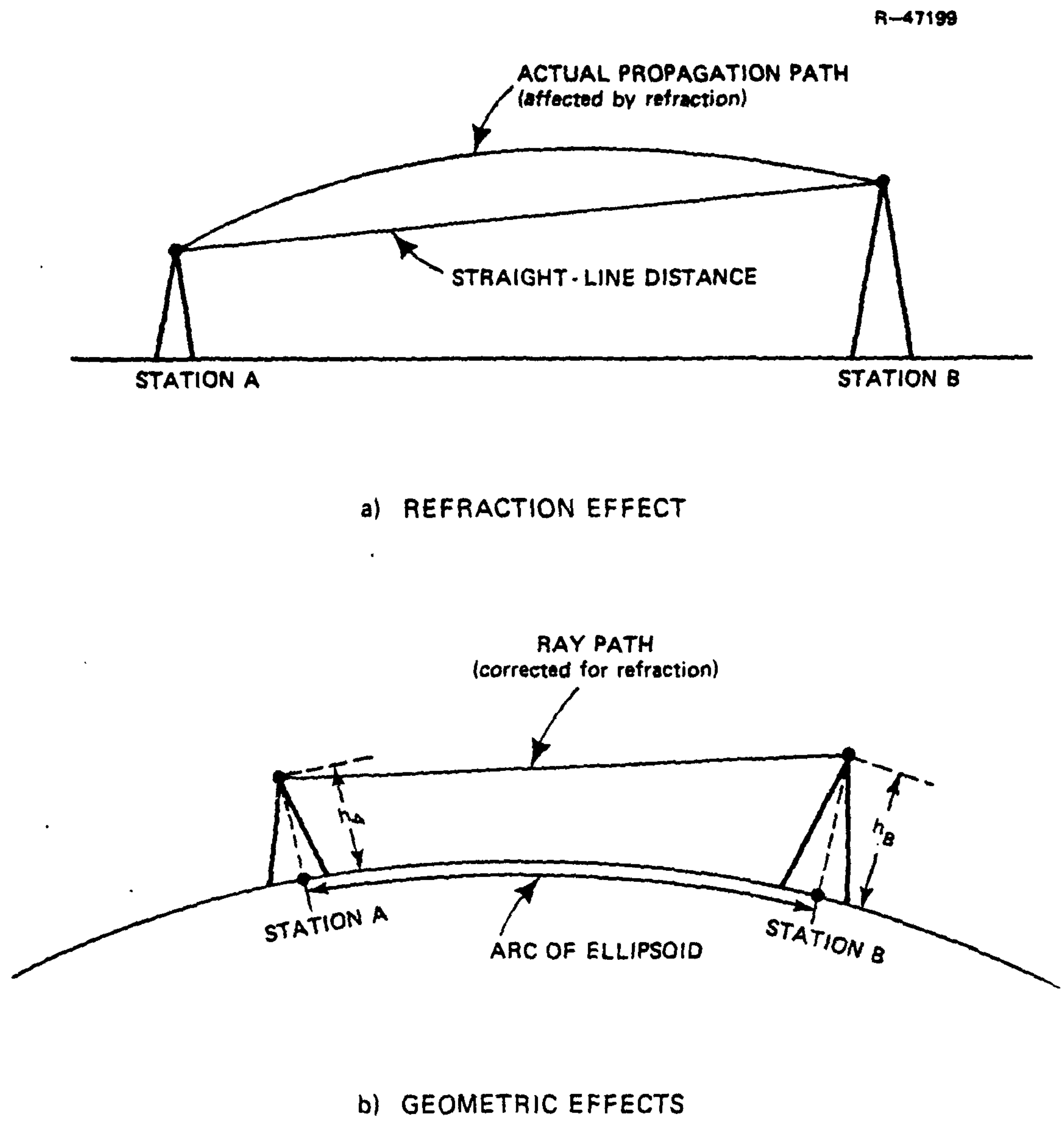


Figure 1.2-8 Refraction and Geometric Effects

Traverse - A third method of establishing the position of survey points is known as traverse. Traverse involves both length and angle measurements, and is analogous to dead reckoning in navigation. A simple example is shown in Fig. 1.2-9a. If the coordinates of the starting point (Point 1) are known accurately, then the measurement of the initial azimuth (α_o), the distances (s_1, s_2, and s_3), and the angles (β_1 and β_2) will permit the computation of the coordinates of Points 2, 3, and 4. In an open traverse like that shown in Fig. 1.2-9a, the accuracy degrades steadily as additional points are added. A method of maintaining accuracy control is to close the loop -- that is, to return to the starting point (Fig. 1.2-9b). The coordinates of the starting point as computed from the traverse (that is, by computation from the coordinates of Point 5, using distance s_5 and angle β_4) are compared with the known values for these coordinates. The discrepancies provide a measure of the accuracy level maintained in the traverse process. This procedure, known as closure, is used in all precise survey methods for accuracy control. Figure 1.2-9b illustrates the concept of a closed (or polygon) traverse.

A more accurate approach to position determination by traverse is illustrated in Fig. 1.2-9c. Here there are two known points (Points 1 and 5 in Fig. 1.2-9c), and the computations can proceed from the ends toward the middle, thus reducing the opportunity for error buildup.

R—50625

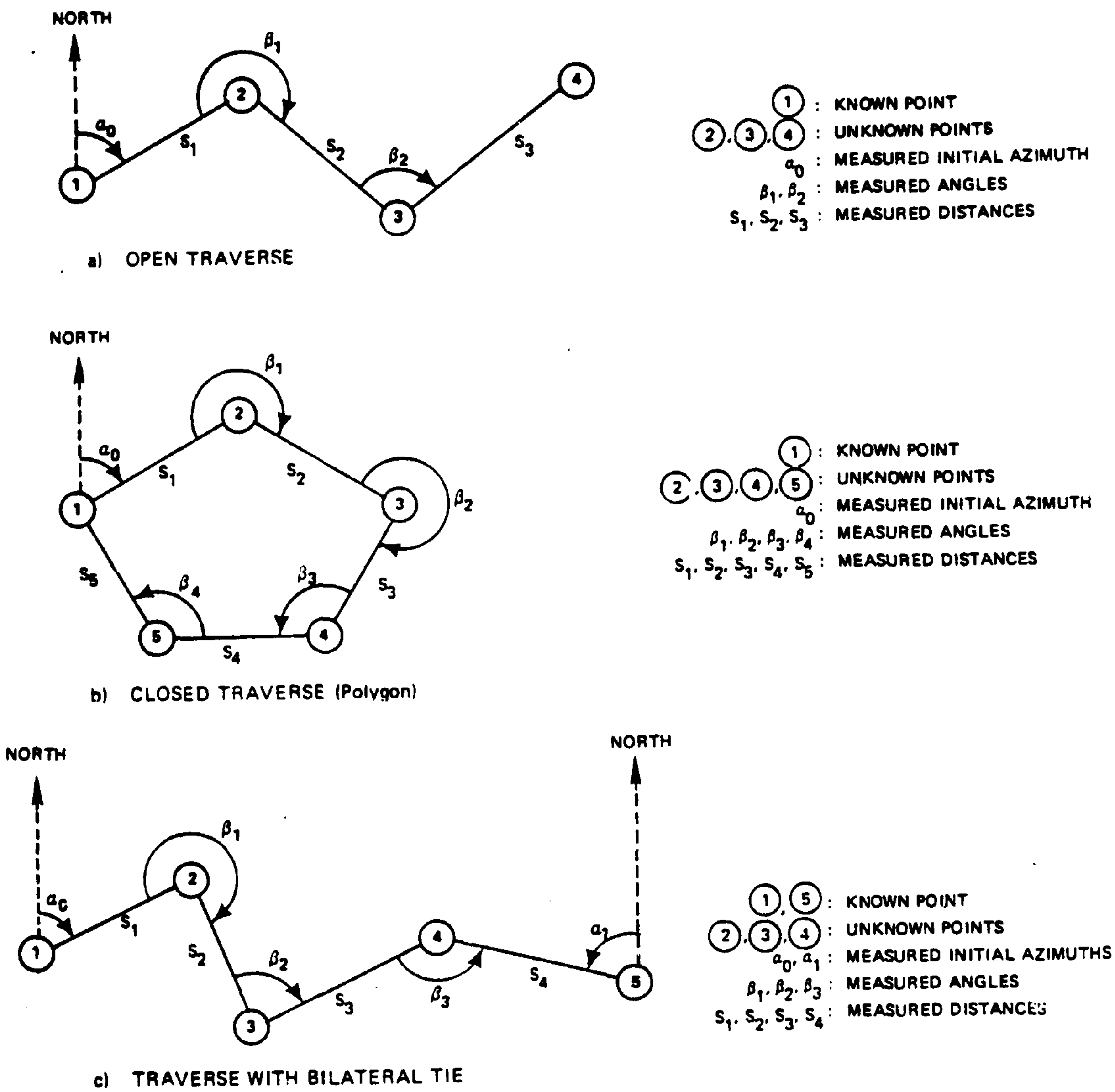


Figure 1.2-9 Examples of Traverse

Vertical Control - Triangulation, trilateration, and traverse are methods of horizontal control -- that is, determining the geodetic latitude and longitude of various points on the surface of the earth. It is also necessary to provide vertical control, by determining the height of a point above the surface of the ellipsoid (measured along a local normal to the ellipsoid). In general there is no direct way to relate a point to the ellipsoid -- except over short distances by computations based on points whose height with respect to the ellipsoid is known (or assumed). Most methods of vertical control are more directly related to the geoid because:

- They are based on mean sea level
- They use the concepts of level surfaces (surfaces of constant potential) and local vertical (plumb line).

The student should note that even for a point at mean sea level (on the geoid), the height determination is not trivial, because geoid and ellipsoid may deviate by as much as 100 meters. The RMS geoid undulation is about 30 meters. For points not at sea level, two kinds of height are used:

- Orthometric (or normal) height (distance from the geoid measured along a plumb line)
- Geodetic (or ellipsoidal) height (distance from the ellipsoid measured along a normal to the ellipsoid).

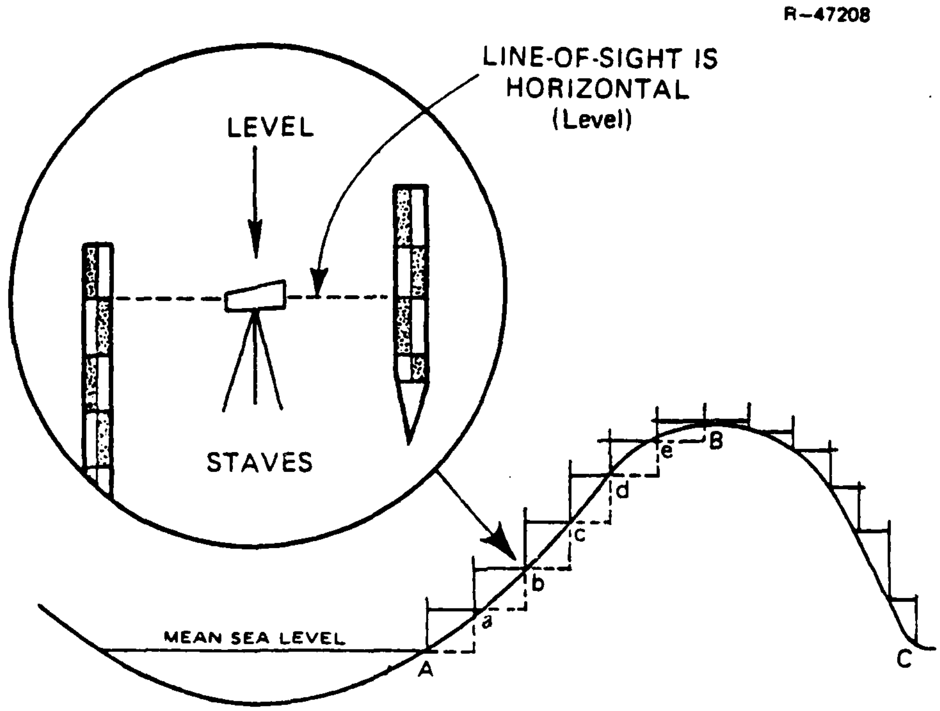
The relation between them is very complicated and depends on the local structure of the gravity field. For certain applications, of course, such as geodetic surveys executed primarily for mapping purposes, there is no problem in the fact that

geodetic positions are referred to an ellipsoid and the elevations of the positions are referred to the geoid. But, in general (especially for such applications as the targeting of missiles), an adjustment in the elevation information to compensate for the undulations of the geoid above and below the regular mathematical surface of the ellipsoid is essential. Such an adjustment uses complex geodetic techniques, requiring precise knowledge of the gravity field. This is an example of the close connection between geometric geodesy (dealing with size, shape, and relative locations) and physical geodesy (dealing with the structure of the gravity field).

There are three commonly used leveling techniques for determining height above sea level -- differential, trigonometric, and barometric -- which yield information at different levels of accuracy. Differential leveling is the most accurate of the three methods. With the leveling instrument (Fig. 1.2-10) locked in position, readings are made on two calibrated staves held in an upright position ahead of and behind the instrument. The difference between readings is the difference in elevation between the points.

The optical instrument used for leveling contains a bubble tube (or spirit level) to adjust it in a position precisely parallel to the geoid. When properly set up at a point, the telescope is locked in a perfectly horizontal (level) position so that it will rotate through a 360° arc. The exact elevation of at least one point in a leveling line must be known; the rest of the elevations are computed from it.

Trigonometric leveling (Fig. 1.2-11) involves measuring a vertical angle from a known distance with a theodolite and computing the elevation of the point. With this method, vertical measurements can be made at the same time horizontal



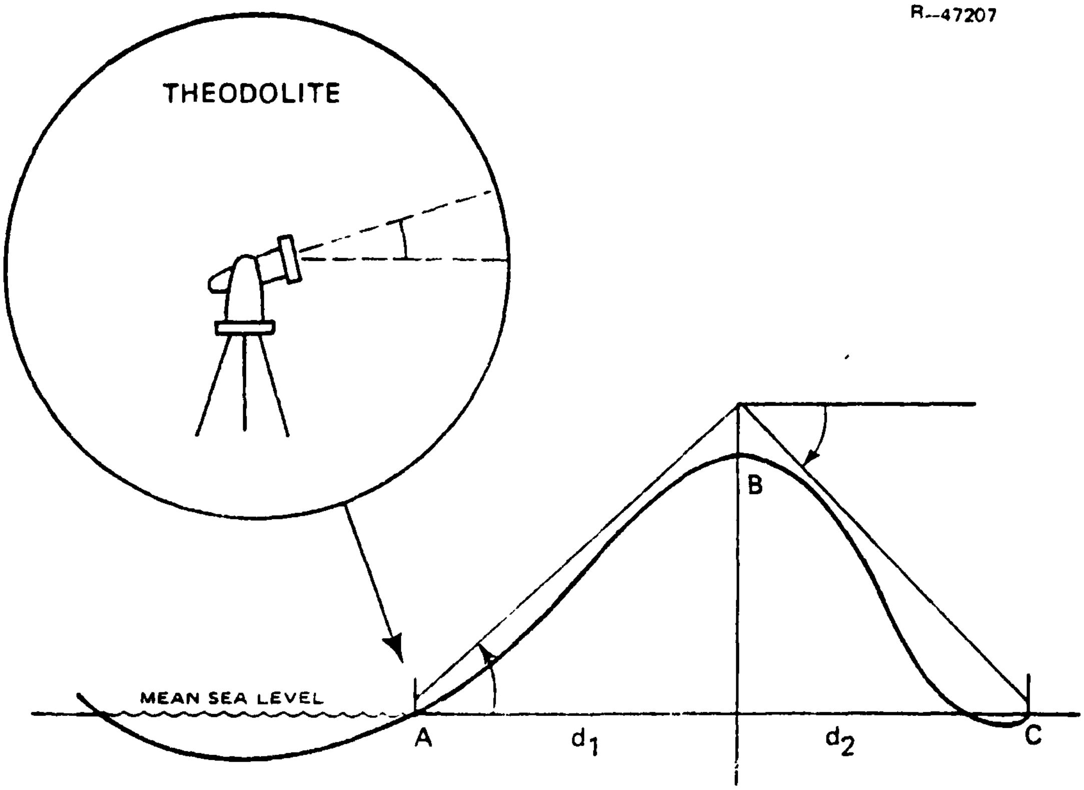
KNOWN DATA:
Elevation of starting point, A.

MEASURED DATA:
Elevation differences, a,b,c,d, etc.

COMPUTED DATA:
Elevation of B, C and all other points.

Figure 1.2-10 Elevation Determination by Differential Leveling

angles are measured for triangulation. It is, therefore, a somewhat more economical method but less accurate than differential leveling. It is often the only practical method of establishing accurate elevation control in mountainous areas.



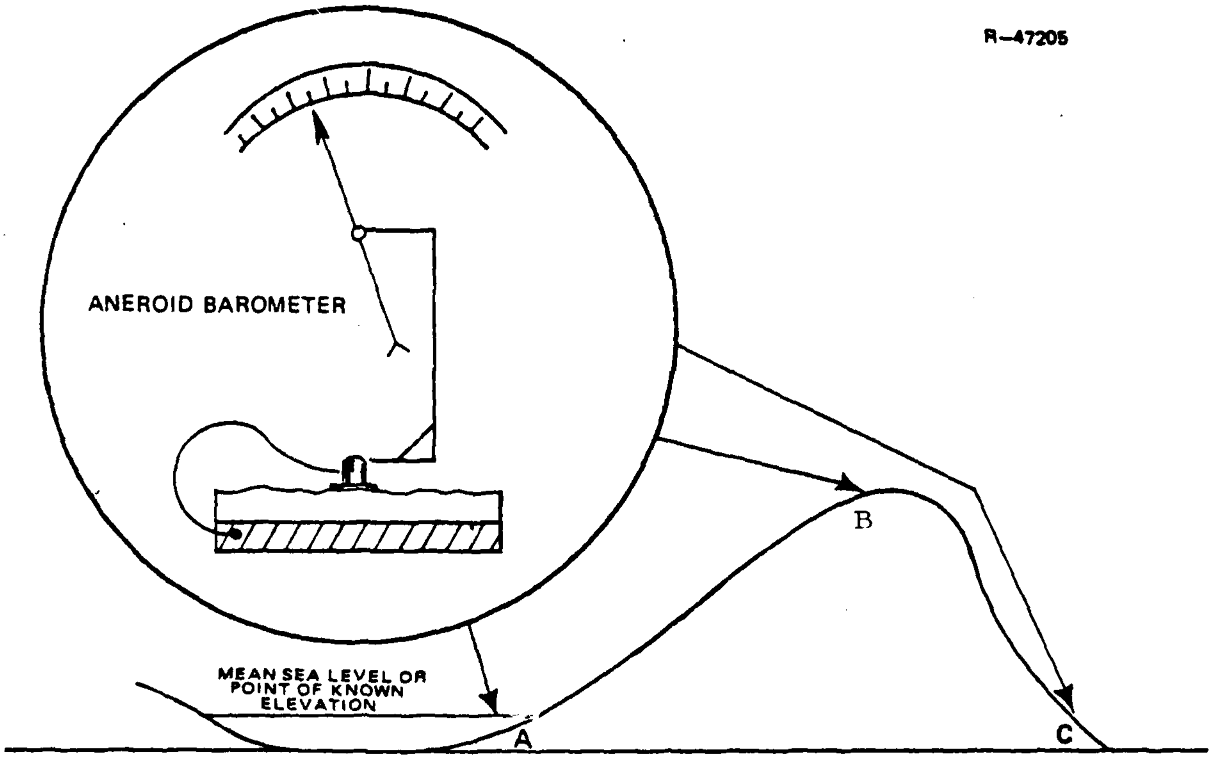
KNOWN DATA:
Elevation of starting point, A.
Horizontal distances, d_1, d_2 between points.

MEASURED DATA:
All vertical angles.

COMPUTED DATA:
Elevation of B, C and all other points.

Figure 1.2-11 Elevation Determination by the Trigonometric Method

In barometric leveling (Fig. 1.2-12), differences in height are determined by measuring the differences in atmospheric pressure at various elevations. Air pressure is measured by using mercury or aneroid barometers, or a boiling point



KNOWN DATA:
Elevation of starting point A.
Meteorological data.

MEASURED DATA:
Air pressure at A, B, C and all other points.

COMPUTED DATA:
Elevation of B, C and all other points.

Figure 1.2-12 Elevation Determination by the Barometric Method

thermometer. Although the degree of accuracy possible with this method is not really suitable for survey work, it can be used to obtain relative heights between widely separated points very rapidly. It is also used in reconnaissance and exploratory surveys when more precise measurements will be made later (or are not required).

All three of these methods are based ultimately on the surface corresponding to mean sea level, determined by obtaining an average of the hourly water heights for an extended period (generally a number of years) at a tidal gauge. As a general surveying strategy, precise geodetic leveling is used to establish a basic network of vertical control points. From these, the height of other positions in the survey can be determined by supplementary methods of lesser accuracy, as required.

1.2.3 Photogrammetry

In the most general sense of the term, photogrammetry refers to the measurement and analysis of photographic images in order to determine the size, shape, and relative configuration of the objects visible in an image. In the context of mapping, charting, and geodesy, the images involved are photographs of the surface of the earth taken from aircraft or orbiting satellites; the goal of photogrammetric techniques is to prepare accurate maps using the information contained in the photograph. This section presents a brief review of the essentials of photogrammetry.

Basic Principles - The question of how to prepare accurate planimetric and/or topographic maps from aerial photographs is the essential material of this section. The answer to this question involves the following elements:

- Aerial photography
- Ground control surveys
- Analytical aerotriangulation
- Stereocompilation
- Final presentation.

Aerial Photography - Precision aerial cameras mounted in fixed-wing aircraft are largely standardized within the commercial photogrammetric industry as well as among those federal agencies having mapping/charting responsibilities. The industry-standard aerial camera yields a 9 in by 9 in negative image on 9.5 in wide aerial film having roll lengths up to 400 ft. It has a nominal focal length of 6 in, and is equipped with a distortion-free lens.

Knowing the focal length of the lens and the flight altitude above mean terrain, a user of aerial photography can determine the scale of any resultant vertically exposed negative by use of the following formula:

$$\text{Scale ratio} = \frac{\text{camera focal length}}{\text{flight altitude}} \qquad (1.2\text{-}3)$$

For example, using the above-mentioned camera (with a 6 in focal-length lens) from an aircraft flying at an altitude of 6000 ft above mean terrain,

$$\text{Approximate scale ratio} = \frac{0.5}{6000} = 1{:}12{,}000 \qquad (1.2\text{-}4)$$

meaning that 1 in on the photographic negative is equal to 12,000 in on the ground.

This simple calculation leads to the next logical step of determining the size of the area covered by the negative. Since the approximate scale is 1 in = 1000 ft, and the negative is 9 in by 9 in, the photographic exposure covers a square 9000 ft on a side. Another simple means of calculating coverage can be derived from the above example by noting the ratio of coverage to flight altitude. It is seen, in this example, that coverage is a factor of 1.5 times the flight altitude. This method of calculating scale must be considered an approximation, for there are, in reality, several factors working against the achievement of true scale in any aerial exposure:

- Variations in aircraft altimeter readings caused by local temperature/barometric changes

- Inability of the camera to achieve a true vertical exposure because of minor aircraft tip/tilt residuals at the moment of exposure

- Atmospheric refraction tending to bend the light-rays entering the camera lens

- Distortions within the camera lens and inability of the camera to hold the film in a true flat plane

- Distortions caused by vertical relief of the terrain being photographed.

Of the above factors, the last one tends to be the most pronounced, especially in terrain having excessive relief. Terrain at higher altitudes appears larger than lower terrain.

Since stereoscopic viewing of aerial photography is essential to the photogrammetric process, it is necessary to extend the concept of single-exposure geometry to multiple-exposure stereoscopic relationships. Figure 1.2-13 shows how

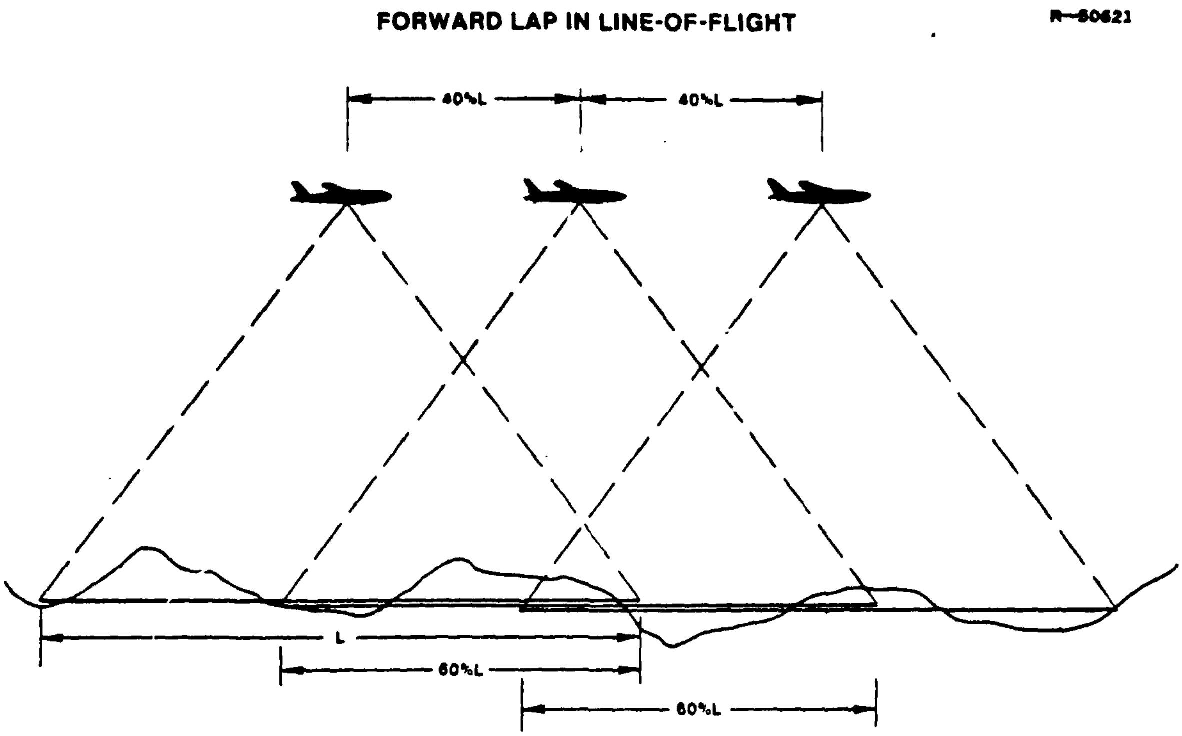


Figure 1.2-13 Stereoscopic Aerial Photography Along a Linear Flight Path

stereoscopic aerial photography along a linear flight path is achieved. In practice, the aircraft tracks along a line on the ground from a specified flight altitude. The cameraman monitors the flight line through a viewfinder and instructs the pilot to make minor left/right heading changes in order to maintain the track. Often, the viewfinder shows a moving grid that the cameraman can synchronize with the moving terrain. With such synchronization achieved, the intervalometer connected with the camera will trigger each exposure at a preset percentage of forward overlap. An approximate 60% forward overlap is standard. If a single flight line of overlapping photographs fails to provide stereoscopic coverage for a particular site or area, it is necessary to fly additional parallel flight lines having side lap relationships of approximately 30% (see Fig. 1.2-14). It should be emphasized that the forward lap

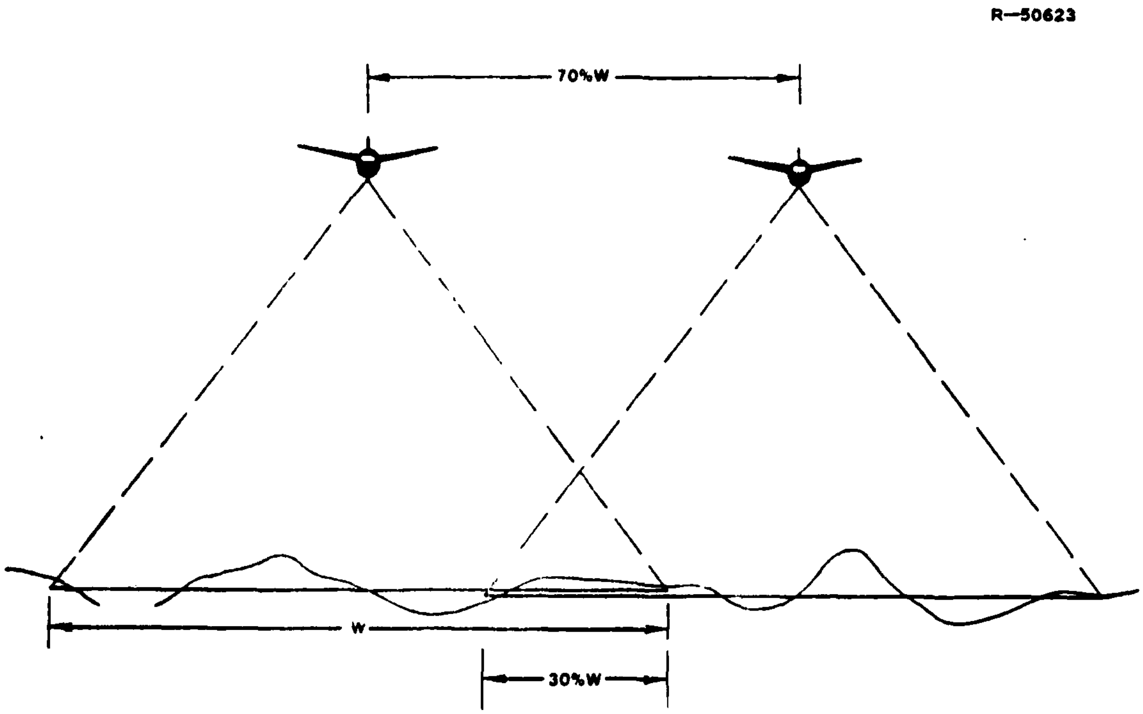


Figure 1.2-14 Side Lap and Spacing of Adjacent Flight Lines

and side lap percentages just mentioned are averages, and are subject to adjustment in coping with extreme differentials in terrain relief.

In addition to the terms forward lap and side lap, there is another term useful in understanding the photogrammetric process: stereo-model (or simply model). In any adjacent pair of exposures taken along a flight line, the gross model is the total area for which the same limits of imagery appear on each exposure. Within the gross model is the neat model, which defines the practical limits for which photogrammetric data are to be acquired (see Fig. 1.2-15).

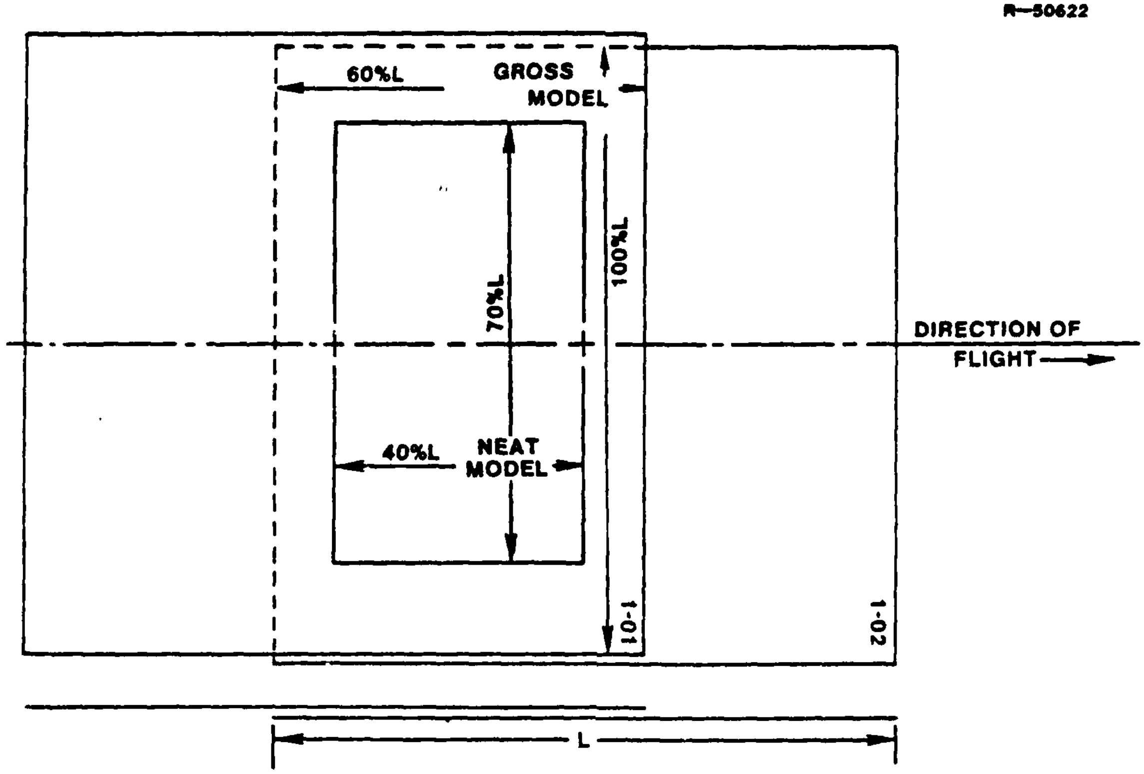


Figure 1.2-15   Computing Size of the Gross Model and the Neat Model

The completion of the aerial photographic phase of a photogrammetric mapping project can yield a wide variety of photographic products, such as 9 in by 9 in contact prints, photo-enlargements, photo-indexes, and glass diapositives (positive contact images on optically flat glass plates). Contact prints are of greatest value to the photogrammetrist because they provide the first stereoscopic view usually available.

Ground Control Surveys - It has already been noted that a single aerial photograph does not possess true scale,

but only an approximation of scale. Likewise, the stereoscopic viewing of a stereo-pair offers two photographs having approximate scale. Since it is the goal of the photogrammetrist to produce a model of the terrain by recreating the precise spatial geometry of the stereo-pair, there is additional information required beyond flight altitude, camera attitude, and distance between exposure stations. The more precise data can be generated through the use of ground control surveys (divided, as indicated in Section 1.2.2, into horizontal control and vertical control).

In carrying out horizontal control surveys, it is the intent to acquire geodetic coordinate values for discrete points (called horizontal picture points) on the ground, which also appear in the aerial photography. In general, these discrete points can be corners of buildings, intersections of fencelines, ends of highway centerline stripes, or any other image point allowing positive identification on the ground of its corresponding imagery in the photography. Such horizontal control surveys are carried out by the methods outlined in Section 1.2.2 -- triangulation, trilateration, and traverse -- or by combinations of these methods.

Vertical control surveys provide the photogrammetrist with elevation data in a similar manner. The vertical control points (vertical picture-points) must provide positive identification on the ground as well as on the aerial photography. Current methods of vertical control include differential leveling, trigonometric leveling, and barometric leveling (as discussed in Section 1.2.2).

Figure 1.2-16 represents typical locations of horizontal/vertical controls within a flight strip. The large outside rectangle indicates the limits of photographic coverage,

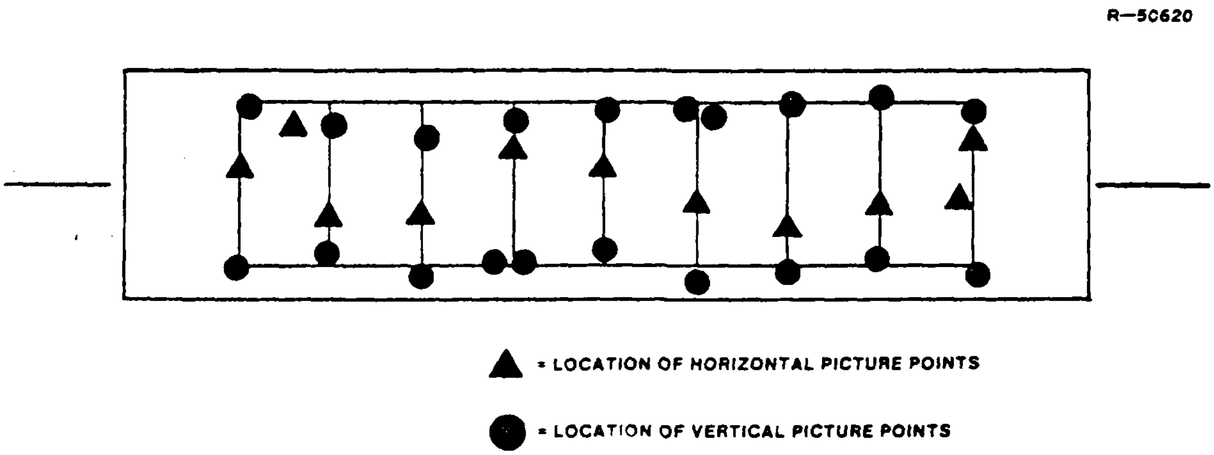


Figure 1.2-16 Sample Flight Strip Showing Best Location for Picture-Point Control

while the small interior rectangles define the individual neat stereo-models within the flight strip. It should be noted that four vertical picture points and a minimum of two horizontal picture points are necessary for each model. Three vertical points can define a datum plane (the fourth point being a test), and two horizontal points allow horizontal scaling. Vertical picture-points are ideally located in the corners of the neat stereo-model, whereas there is no optimum position for horizontal picture points other than the requirement that they form a reasonably long baseline spanning the length of the neat stereo-model. However, ideal conditions seldom occur in practice, and the horizontal/vertical picture-point locations are normally compromised somewhat to fit existing field conditions and accessability to identifiable points.

Standard practice dictates that field-survey personnel pin-prick and label the locations of the horizontal/vertical picture-points on the photographs. The accumulated field data are reduced, computed, and adjusted. Then the final coordinate values are entered adjacent to the pin-pricks for the corresponding picture-points.

Analytical Aerotriangulation - Because of the low density of existing survey control data for a particular project area, or because special situations (for example, rugged terrain) may make it difficult to accomplish control surveys by normal methods, analytical aerotriangulation methods are employed to supplement or substitute for actual field surveys. Referring to Fig. 1.2-16, assume that field surveys can be carried out in only the first, fifth, and eighth stereo-models (reading from left to right). Analytical aerotriangulation methods allow the determination of coordinate values for those picture points in the remaining stereo-models within required accuracy constraints.

Although the details of the subject are beyond the scope of this text, the steps are generally as follows:

- Picture points are selected for those stereo-models that cannot be field-surveyed
- On glass diapositives, small holes are drilled in the photographic emulsion at the identical picture-point locations previously pin-pricked on the aerial contact print
- Each glass diapositive is placed in an optical comparator, and the cartesian coordinates of all picture-points are recorded (to the nearest micron). In addition to the cartesian coordinates of picture-points on each glass diapositive,

the coordinate values for the camera fiducial mark (normally there are four or eight fiducial marks in each photograph) are included as a means of providing a reference frame

- Finally, all the known data (Cartesian coordinates for fiducial marks and surveyed/unsurveyed picture-points, plus field-surveyed coordinate values) are manipulated by computer to yield coordinates for all picture-points.

Stereocompilation - Stereocompilation instruments (or stereoplotters) are used in the final stage of map preparation. This section considers the use of the direct-viewing optical-train stereoplotter. The direct-viewing stereoplotting instrument allows an operator to view a stereo-pair of photographs through a series of prisms in such a way that the left eye is directed to a portion of the left photograph and the right eye to the corresponding portion of the right photograph in the stereo-pair. Within the optical train associated with the eyepieces is a dot, cross, or small circle which serves as a reference mark within the operator's field of view. When the operator first sets up a stereo-model, he must bring the imagery of the left eyepiece into coincidence with the imagery in the right eyepiece through the use of rotational and translational controls. When he has eliminated the parallax (or imagery mismatch) in the stereo-model, the operator has completed the operation called inner-orientation. The next step, exterior orientation, adjusts the stereo-model to its correct scale and level datum.

The reference mark (often termed floating point) that the operator sees in each eyepiece can be displaced along the axis connecting the centers of the two photographs. Movement of a reference mark introduces a parallax which causes an apparent vertical displacement of the reference mark when viewed

against the stereoscopic terrain image. The controls causing this displacement allow direct read-out of the elevation of the reference mark. The reference mark can be made to appear to float above, on, or below the apparent surface of the terrain. Through mechanical linkages the operator can shift the reference mark over the entire stereomodel (in the x and y plane) and can move the reference mark vertically (in the z axis). The same linkages connect to a pantograph, to which a pencil or pen is attached for marking points and lines on the map being prepared.

In practice, the operator plots a building (or any other man-made or natural feature) by bringing the reference mark to the proper elevation of the object, then moving it in the x-y plane around the visual outline of the object. As this is done, the pencil at the end of the pantograph traces a replica of the object. In contract, to plot a contour, the operator sets the elevation of the desired contour on the dial of the z-motion control, moves the reference mark until it is in apparent contact with the surface of the terrain, drops the pencil/pen on the pantograph into contact with the drawing, and finally moves the reference mark in the x-y plane in such a manner as to maintain visual contact with the surface of the terrain. The operator follows the general procedure of first plotting all planimetric features (buildings, roads, fences, cross-country utility lines, vegetation outlines, etc.) then completes the stereo-model by adding contours, spot elevations, or any other specialized data (cross-sections, profiles, etc.). The product of the operator's efforts is a working drawing (in pencil or ink on mylar film to insure that dimensional stability is maintained throughout the project).

Final presentation - Inked or negative-engraved final sheet layouts are made by drafting personnel, who also edit

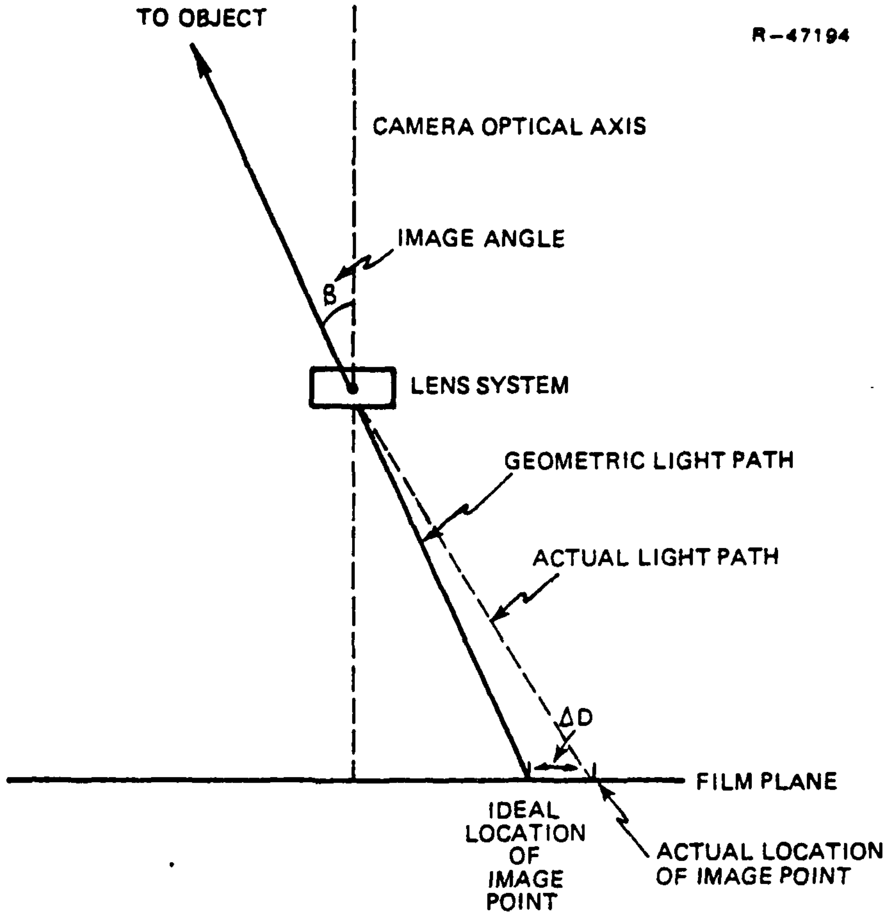
the completed stereo-compiled manuscripts for accuracy and add place names, symbolization, etc. In mapping large areas it is rare for the final sheet layout (normally oriented with North to the top of the sheet) to coincide with the direction of the flight lines. Therefore, finalization of drafting may find any one final sheet extending beyond the limits of one stereo-model into another (either on the same flight line or on an adjacent one).

As just described, the final presentation is in the form of a map or chart. It is also possible to finalize photogrammetric data as <u>digital</u> information for storage and use in a computer. In the case of aerial photogrammetry, such digital data would be referred to as a <u>digital terrain model</u>.

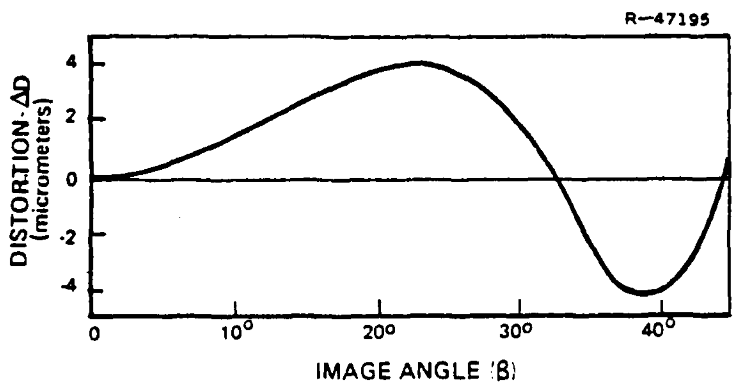
<u>Equipment and Film Considerations</u> - Air (or space) to ground photography serves one of two purposes -- photoreconnaissance or photomapping --and the equipment considerations are considerably different in these two applications. In photoreconnaissance the identification of objects is the most important requirement, leading to the consideration of camera and film resolution as an overriding criterion. For photomapping applications, on the other hand, adequate resolution is certainly important, but the overriding metric considerations are such things as:

- Film flatness
- Stability and rigidity of the camera body and mount
- Minimization of lens distortions.

Not only are lens systems for use in photomapping designed and constructed with great care, and thoroughly tested on an individual basis, but they are generally also provided with detailed



a) GEOMETRY OF OBJECT, LENS, AND IMAGE



b) LENS DISTORTION CURVE

Figure 1.2-17 A Typical Lens Distortion Curve

calibration and residual distortion data, which can be used for correction purposes during the analysis and plotting process. As an example, a hypothetical distortion curve is shown in Fig. 1.2-17. In this plot, the independent variable, β, is the angle between the center of the field (camera axis) and the object. The dependent variable is the linear displacement of the actual image point on the film plane from its theoretical (or geometric) location.

The film also plays an important part in the accuracy of the photogrammetric process. While a discussion of film characteristics is given in Unit Three (Section 3.2.1), it should be emphasized that a major criterion for film used in photomapping is dimensional stability. This depends not only on the film base material but also on the details of photographic processing and on storage conditions. Ideally, it would be best to use photographic materials coated on glass plates, but the obvious inconvenience in handling and storing such materials, as well as the increased weight, often precludes this choice. The dimensional stability is an important consideration not only for the photographic film on which exposures are made in an aircraft or a satellite, but also with regard to the print film or photographic paper on which images are printed for use by photo-interpretation personnel. Ordinary photographic paper, for example, undergoes dimensional changes (non-uniform in direction) of as much as three percent during the various stages of processing and drying.

1.2.4 Geodetic Astronomy

Because observations of stars permit the determination of directions that can be regarded, for practical purposes, as <u>fixed in space</u> (and, hence, unaffected by the rotation of the earth and its motion through space), there is a

branch of geodesy that is concerned with the use of astronomic techniques in determining the location of points on or near the earth's surface. As a preliminary to the discussion of astronomic positioning, it is necessary to review two topics of an astronomical nature:

- The rotational and translational motions of the earth, as observed in a coordinate system fixed with respect to the distant stars -- a so-called inertial* coordinate system (Section 1.2.4.1). This topic also includes a brief summary of the astronomic determination of time.

- The establishment and use of inertial coordinate systems, related to star positions, that are important in astronomy and geodesy (Section 1.2.4.2).

1.2.4.1 Rotation and Translation of the Earth with Respect to Space

The motion of the earth with respect to a coordinate system fixed in space involves:

- The translational motion of the center of mass through space

- The rotational motion of the earth around its center of mass.

The inertial coordinate system, fixed in space, with respect to which earth motion is defined, is related to the apparent positions of the distant stars and is usually called the stellar inertial reference frame. This coordinate system is discussed in Section 1.2.4.2.

---

*The student will recall from physics and mechanics courses that an inertial coordinate system is one in which Newton's laws of motion are valid. An alternative characterization: a coordinate system which is non-rotating and not accelerating.

Relating positions on the earth to a frame fixed in space is required, for example, in celestial navigation and in the observation of artificial satellites. In either of these cases, observations are made from a point on or near the earth's surface (whose position is defined in terms of a coordinate system moving with and rotating with the earth) of objects whose position is defined in terms of, or computed in, the space-fixed inertial frame.

The translational motion of the earth is easily described in terms of the earth's orbital motion around the sun. For more precise purposes, it is better to consider the motion of the center of mass of the earth-moon system around the center of mass of the solar system (which is very close to, but not coincident with, the geometrical center of the sun); this is supplemented by a description of the motion of the center of the earth around the center of mass of the earth-moon system. In general terms, the motion of the earth is elliptical with the sun at one focus. In addition it undergoes slight perturbations caused by the other planets in the solar system (particularly Jupiter). The plane of the earth's orbit is known as the ecliptic. Depending on the degree of accuracy required, various sets of formulas and tabular representations are available to predict the position of the earth with respect to other bodies in the solar system as a function of time.

It is more difficult to describe the rotational motion of the earth around its center of mass. Such a description must consider these two aspects:

- The orientation of the earth's spin axis with regard to the axes of the inertial space-fixed coordinate frame

- The orientation of identifiable features on the crust of the earth with respect to the spin axis.

Both aspects of the rotational motion of the earth will be discussed in the present section.

The Coordinate Systems - The two fundamental planes for the definition of stellar and earth-fixed coordinate systems are the ecliptic (the plane of the earth's orbit around the sun) and the equator (the plane normal to the earth's spin axis, passing through the center of the reference ellipsoid), as shown in Fig. 1.2-18. These planes intersect in a line called the nodal line or the line of equinoxes. This line marks the apparent direction of the sun at the vernal (Spring) and autumnal (Fall) equinoxes, when the sun appears to be overhead at the equator, and the day and night are of equal length. The vernal equinox is taken as the fundamental direction (x-axis) for the space-fixed system. The plane of the equator is the x-y plane, and the z-axis is parallel to the earth's spin axis.

For the earth-fixed reference frame, the z-axis is also taken to be the earth's spin axis and the x-y plane is the equator. The x-axis is defined to be the direction from the center of the earth to the meridian of Greenwich, with the y-axis completing a right-handed coordinate system. Because of the non-rigidity of the earth and the existence of significant crustal motion, this definition is not complete. Further details are considered later under the heading Polar Motion.

Rotation - Because of the rotation of the earth, the angle between the x-axis of the earth-fixed system and the x-axis of the space-fixed system advances through 360 deg every day. This angle, called the Greenwich sidereal time, is the basis for ordinary time keeping, and determines Universal Time

R—50624

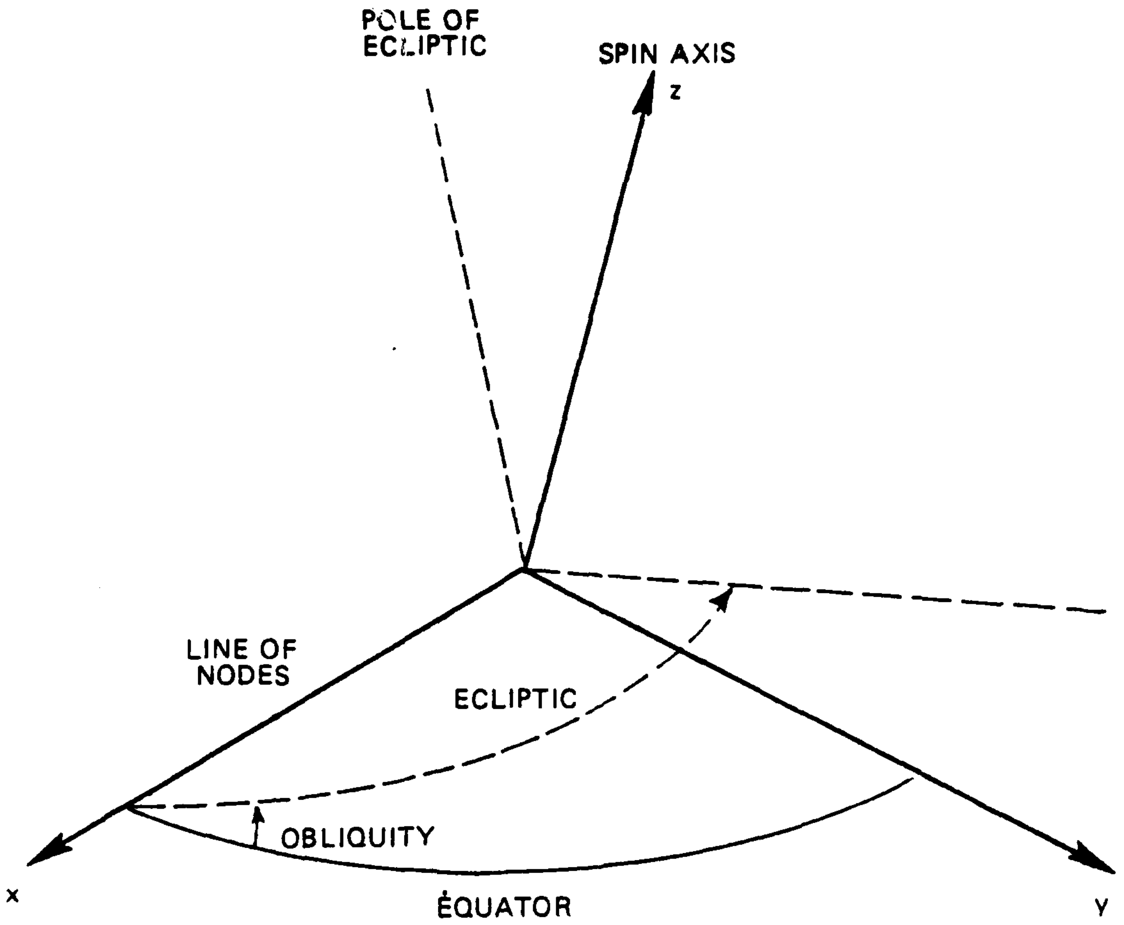


Figure 1.2-18 Space-Fixed and Earth-Fixed Coordinate Systems

(UT) and <u>civil time</u> (standard time for the various time zones). It is customary to express this angle in time-like units, rather than in degrees. Thus, one complete revolution is equivalent to twenty-four hours; one quarter of a revolution (90 deg) is six hours; etc. Although historically it was assumed that the earth's rotation rate was constant, it has been discovered in the present century that this is not the case. The earth is subject to a gradual deceleration caused by the friction of the tides, particularly in shallow seas and along the continental margins. This tidal deceleration causes an increase in the length of the day by approximately 4 msec per century. In addition, there are seasonal variations in the rotational speed of the earth, with periods of a year and a half-year.

Superimposed on these are a large number of smaller effects, some of which appear to be random. The relation between the earth's rotational orientation (i.e. Universal Time) and the uniform time as kept, for example, by an atomic clock is tabulated in astronomical references; however, it is not possible to make precise predictions of the earth's rotational orientation at times far in the future. Further details of the timekeeping aspects of the earth's rotation are discussed in a later section.

Motion of the spin axis - It has been known since ancient times that the spin axis of the earth is not fixed in space. The present North Star (Polaris) -- less than a degree from the actual pole right now -- was nowhere near the pole in 3000 BC, when a different star (now called α Draconis) was the pole star. The motion of the pole has a long-period component known as precession, superimposed upon which is a short-period component of much smaller amplitude, the nutation. These two motions will now be discussed in detail.

Precession - Precession refers to the very long-period motion of the axis of rotation of the earth with respect to inertial space. The effect is caused by the gravitational torques of the sun and the moon and is similar in nature to the precession of a spinning top in the presence of a gravitational field. Historically this phenomenon was detected as early as the second century BC, when it was noticed that the observed longitudes of stars show a gradual change with time. This systematic motion can be construed as a secular (changing steadily with time) motion of the vernal equinox.

The gravitational torques of the sun and the moon cause the earth's spin axis to describe a circular motion in the sky about the pole of the ecliptic. In addition, the

gravitational effects of the other planets (principally Jupiter) cause small changes in the orientation of the earth's orbital plane; hence, the ecliptic also undergoes a regular motion with respect to an inertial coordinate system. As a consequence, the vernal equinox (defined by the intersection between the equatorial and ecliptic planes) moves in space because both of these planes change their orientation with the passage of time. The motion of the vernal equinox, and of the stellar reference frame for which it is the principal direction, is known as <u>general precession</u>. The two contributions to general precession are the motion of the equatorial plane, called <u>luni-solar precession</u>, and the motion of the ecliptic plane, called <u>planetary precession</u>. The spin axis describes a circular motion in space with a period of twenty-five thousand eight hundred years. The radius of the motion is approximately 23.5 deg.

<u>Nutation</u> - Precession, as described above, refers to the long-period circular motion of the earth's axis of rotation. Superimposed on this motion are a variety of smaller periodic motions in space. This set of motions is called <u>nutation</u>. As a result of nutation, the actual axis of rotation describes a complicated series of more or less elliptic motions about the average position described by the theory of precession.

The nutational motion is resolved into two components. The first, called the <u>nutation in longitude</u>, is the motion of the vernal equinox along the plane of the ecliptic. The second component, called the <u>nutation in obliquity</u>, is the variation in the angle between the equatorial plane and the plane of the ecliptic. The periods of these nutational motions are much shorter than the precessional motion, ranging from 18.6 years down to about 5 days. The principal components of the nutational motion, with their periods and amplitudes, are listed in Table 1.2-3.

TABLE 1.2-3

COMPONENTS OF THE NUTATION

COMPONENT	PERIOD	AMPLITUDE ($\widehat{sec}$)
Principal nutation	18.6 years	9.21
Semi-annual nutation	0.5 years	0.57
9-year nutation	9.0 years	0.09
Semi-monthly nutation	13.7 days	0.09
Annual nutation	1.0 years	0.06
Monthly nutation	27.6 days	0.02
122-day nutation	122.0 days	0.02

Polar Motion - The theory of precession and nutation describes the motion of the earth's spin axis with respect to a reference frame fixed in space. But the axis of rotation is not fixed with respect to the surface of the earth. The changing orientation of the axis of rotation with respect to an axis fixed in the body of the earth is called polar motion. It is made up of a number of components that must be determined observationally. A rigorous mathematical theory for the polar motion has not been developed.

The major component of the polar motion is known as the free nutation or the Chandler wobble. The period of this motion, which cannot be accounted for exactly without considering the elasticity of the earth, is approximately 430 days. The excitation and damping mechanisms of the free nutation are poorly understood and cannot be modeled precisely. In addition to the 430 day free nutation, there is also an annual motion of the axis of rotation. The amplitude of this motion is comparable to that of the free nutation (about 0.1 $\widehat{sec}$). Recent

research indicates a strong correlation of the annual component with the redistribution of mass in the atmosphere associated with meteorological phenomena.

Local Effects - In addition to the global aspects of the motion of an earth-fixed reference frame in a space-fixed coordinate system, local motions must also be considered for extremely accurate work. These include the effects of solid-earth tides* and both local and large-scale deformations of the crust (plate tectonics). The magnitude and direction of these motions are strongly dependent on local geology and require the location of geophysical instruments at the observation site to evaluate any effects that might occur. Generally, these are phenomena of small magnitude.

Time Based on Earth Rotation - The rotational speed of the earth is not observed directly. Rather, the rotational rate is determined from observations of the angle of rotation over a period of time. The most important information for navigation as well as geodesy is the relationship between these two time scales:

- A uniform time scale derived from atomic clocks
- A time scale defined in terms of the rotation of the earth.

Universal Time (UT) is the earth rotation time scale used as the basis for time keeping all over the world. UT is related to the actual rotation of the earth with respect to a

*That is, motion of the earth's crust due to the gravitational attraction of the sun and moon.

space-fixed reference frame. In practice, it is determined from astronomical observations of stars. Numerous observatories all over the world include, as part of their observing program, the measurement of star transits across the meridian for the purpose of maintaining the precise calibration of Universal Time.

The time determined directly from star observations is designated UT0. UT0 refers to the instantaneous orientation of the pole of rotation with respect to the surface of the earth. If corrections are applied for polar motion, in order to reorient the measurements to an internationally agreed upon mean polar position,* the resulting time scale is known as UT1. Corrections for the seasonal variation in rotational speed, determined empirically from past observations, are applied to UT1 to arrive at a third time scale, known as UT2. Each of these time scales is directly related to the rotation of the earth and is, therefore, not precisely uniform.

For scientific purposes, accurate time scales based on the use of atomic clocks of various kinds (cesium clock, rubidium clock, hydrogen maser) have been developed within the last decade. <u>Atomic time</u>, as kept by such clocks, has been coordinated with another physical time scale called <u>ephemeris time</u>, derived from a study of the motion of the planets and other bodies in the solar system (primarily the motion of the moon around the earth). Ephemeris time is regarded as the independent variable in the equations of motion that define the physics of the solar system. Considerable effort has been expended to achieve consistency between the atomic and the ephemeris scales.

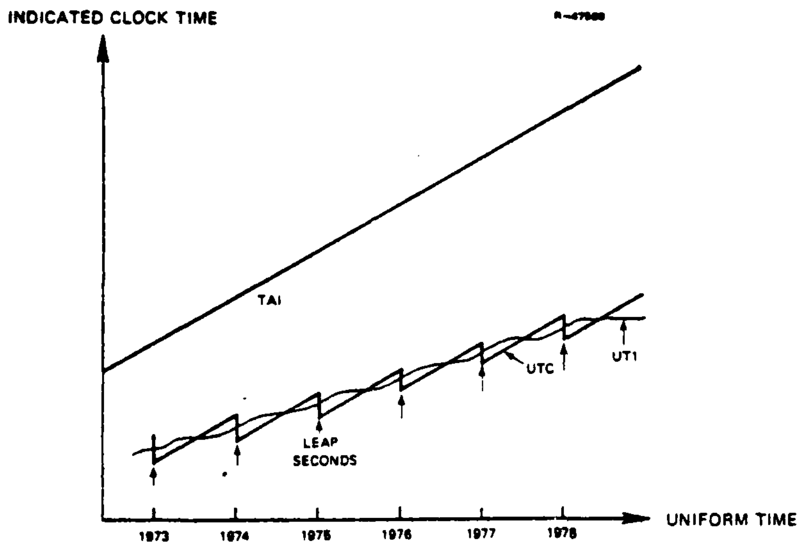
---

*Known as the Conventional International Origin (CIO).

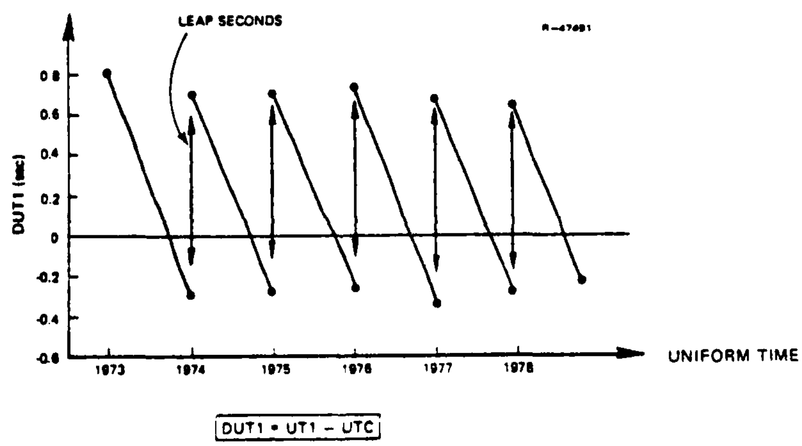
Despite the advantages of atomic time for precise scientific work, it is still necessary to maintain a time scale that will be consistent with the irregular rotation of the earth, because time is used, in navigation and in astronomy, to describe the rotational orientation of an earth-fixed frame with respect to the space-fixed coordinate system. Thus, a perfect atomic clock would not, after a lapse of many years, correctly predict such phenomena as sunrise, sunset, star transits, eclipses, etc. A time scale has been adopted internationally which attempts to combine the best features of atomic time with the earth-synchronous requirement of Universal Time. This scale, known as Coordinated Universal Time (UTC), is now the basis for civil time keeping and is broadcast by time and frequency standard radio stations throughout the world (in the United States, the National Bureau of Standards' WWV and related stations).

UTC is a uniform time scale in that it is based on atomic time. It is closely compared with the Universal Time scale based on earth rotation, and whenever the two scales have deviated by an appreciable fraction of a second, a step adjustment by one second is made in the UTC clock. Such an adjustment is known as a leap second. At the present time, leap seconds are necessary once or twice a year. Whenever feasible, the adjustments are made at the end of December and the end of June. The goal of the leap second adjustments is to keep the difference between UTC and UT1 from ever exceeding 0.8 seconds. Along with UTC, the time and frequency standard broadcasts make available the predicted (or estimated) difference between UT1 and UTC. This difference, known as DUT1, permits the user to calculate the earth rotation time. On the other hand, the transformation from UTC to the International Atomic Time scale is simple, since the two scales differ by an integral number of seconds (equal to the number of times the leap second correction has been applied).

Figure 1.2-19 is a graphical representation of the relationships among the International Atomic Time (TAI), UTC, and UT1 scales during the period 1973 to 1978.



a) THE THREE TIME SCALES



b) THE VARIATION OF DUT1

Figure 1.2-19 The TAI, UTC, and UT1 Time Scales

Observational Techniques - A number of techniques exist for the determination of the orientation of the earth with respect to a non-rotating space-fixed coordinate system, generally based on observational data obtained from precise optical instruments. As an example, the transit circle has been used for precise determination of the positions of stars and planets, in order to determine the constants describing precession and nutation. The transit circle is an instrument consisting of a refracting telescope that can rotate about a fixed horizontal axis precisely oriented in an East-West direction. The telescope is thus constrained to move only in the plane of the meridian.

Observations of stars for the purpose of determining polar motion and time are generally made with the following precise optical instruments:

- The visual zenith* telescope (VZT)
- The photographic zenith tube (PZT)
- The Danjon astrolabe.

The VZT is designed to be oriented in the plane of the local meridian to measure zenith distances of stars as they transit the meridian. The zenith is defined by spirit levels attached to the instrument. Measurements are made with a micrometer at the eyepiece of the telescope. Such instruments have been used since 1900 to observe up to 10 stars per night at various locations to determine polar motion (also called the variation of latitude). The PZT is a long-focus refracting telescope constrained to point only in the direction of the zenith. A shallow basin of mercury reflects the incident light from the

*Zenith refers to the upward vertical direction.

lens to a photographic plate. The surface of the liquid mercury is parallel to the local geopotential surface; hence, the perpendicular to the mercury surface defines the direction of the local zenith. A typical night's observing program involves four observations of about thirty stars to produce an estimate of time as well as latitude variation. The Danjon astrolabe is an optical instrument involving visual observation rather than the use of a photographic plate.

Another source of polar motion data is Doppler measurements of navigation satellites. The satellite orbit is used as a reference in space, and range differences to the satellite with respect to time from each of a number of tracking stations are used to determine the polar motion components of the earth's orientation with respect to the satellite orbits. Other techniques, such as laser ranging to the moon and radio interferometry, are expected ultimately to provide higher accuracy than the classical optical techniques for the determination of polar motion and Universal Time.

International Distribution of Time and Polar Motion Data - In order to predict the precise orientation of the earth's spin axis and the precise orientation of an earth-fixed reference frame with respect to the space-fixed coordinate system, it is necessary to have up-to-date information regarding polar motion and the Universal Time correction. Such information is collected and distributed internationally by the cooperative effort of a number of countries. Polar motion information is contributed by the International Polar Motion Service (IPMS). Time information, as well as polar motion data, is the responsibility of the Bureau International de l'Heure (BIH). Cooperating organizations within the United States for polar motion and time determination include the U.S. Naval Observatory, the Defense Mapping Agency, and the National Geodetic Survey (now part of the National Ocean Survey).

A major source of polar motion data has been the five International Latitude Stations, the locations of which are listed in Table 1.2-4. These stations are equipped with visual zenith telescopes and determine the instantaneous position of the pole with respect to a mean pole position called the Conventional International Origin (CIO). Polar motion is determined by the use of least-squares procedures to analyze the variation of latitude at each of the five stations for the common component due to polar motion. The results appear in a monthly report issued by the International Polar Motion Service (IPMS), which also receives and processes data from a second network of fifty-four stations.

TABLE 1.2-4
THE INTERNATIONAL LATITUDE STATIONS

STATION NUMBER	LOCATION
1	Gaithersburg, MD, USA
2	Ukiah, CA, USA
3	Mizusawa, Japan
4	Kitab, USSR
5	Carloforte, Italy

The BIH uses the results obtained from observational programs involving approximately 70 astronomical instruments, together with data from the Doppler tracking of satellites, in order to determine both polar motion and Universal Time. Instruments reporting to the IPMS also contribute data to the BIH. Elaborate computational procedures are followed by the BIH to maintain the greatest possible degree of accuracy. Five-day values for the coordinates of the pole and the difference between UT1 and UTC are distributed on a monthly basis.

This information is redistributed in the United States by the U.S. Naval Observatory. The BIH also provides a weekly teletype service listing of coordinates of the pole and UT1 - UTC for users with an immediate need for the data.

Within the United States, the U.S. Naval Observatory makes astronomical observations using photographic zenith tubes located at Washington, D.C. and Richmond, Florida, for the determination of Universal Time and the variation of latitude. Published results are summarized in Table 1.2-5. The National Geodetic Survey is responsible for the operation of the International Latitude Stations within the United States. Their data are contributed to the IPMS, which publishes the results.

TABLE 1.2-5
NAVAL OBSERVATORY PUBLICATIONS

PUBLICATION	FREQUENCY	CONTENTS
Time Service Announcement Series 6	Monthly	Data on the UT1 scale
Time Service Announcement Series 7	Weekly	Coordinates of the pole; DUT1
Time Service Announcement Series 11	Yearly	Summary of Naval Observatory observations and results
Time Service Announcement Series 15	Monthly	Five-day coordinates of the pole and values of DUT1 as determined by the BIH

1.2.4.2 Stellar Reference System

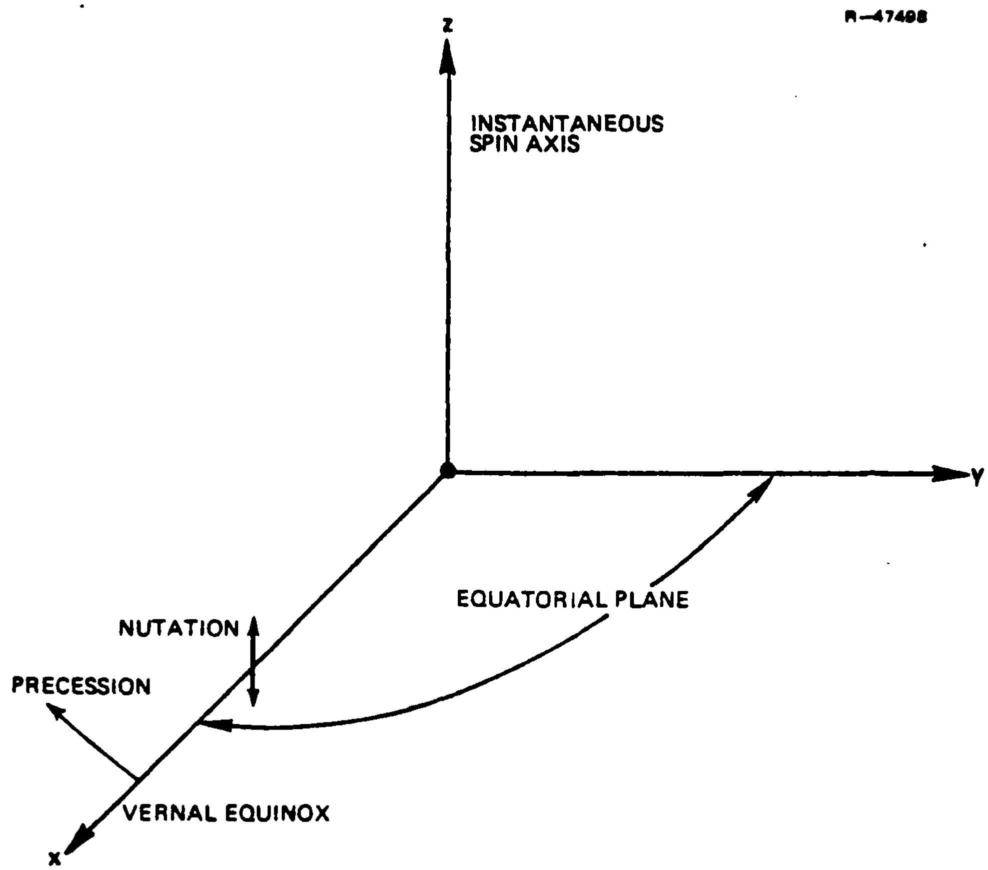
The use of space-fixed coordinate systems to define the motion of the earth has already been referred to in Section 1.2.4.1. Such coordinate systems are also useful for orbital computations (planets and natural satellites, as well as artificial satellites), because the equations of motion are simplest when expressed in an inertial reference frame.

The origin of a stellar inertial reference frame may be chosen for convenience in one of several ways. For example:

- The center of mass of the earth
- The center of mass of the earth-moon system
- The center of mass of the sun
- The center of mass of the solar system.

For the present discussion the origin is taken as the center of mass of the earth.

The directions of the coordinate axes may be defined in several ways, giving rise to a variety of reference frames. For example, the instantaneous direction of the earth's spin axis may be taken as the z-axis. The x-y plane, normal to this, is the instantaneous equator. The x-axis is the vernal equinox of instant (that is, at the moment for which the definition is valid), defined by the intersection of the equatorial and ecliptic planes (see Fig. 1.2-20). Coordinates measured with respect to this frame are identified by the designation true of instant or true of date.



AXES SHIFT WITH RESPECT TO FIXED DIRECTIONS IN SPACE

Figure 1.2-20 Coordinate Frames: True of Instant

The true of date coordinate system just described is not actually space-fixed (or inertial), because of the effects of precession and nutation. When a non-rotating system is required, it can be defined by fixing the coordinate axes in space as they were at some instant of time -- the beginning of the current year, or the beginning of a standard year like 1950 or 1975 (Fig. 1.2-21). The z-axis will then no longer coincide with the actual spin axis, nor the x-y plane with the actual equator. Coordinates measured in such a frame are identified by the date associated with the reference frame -- for

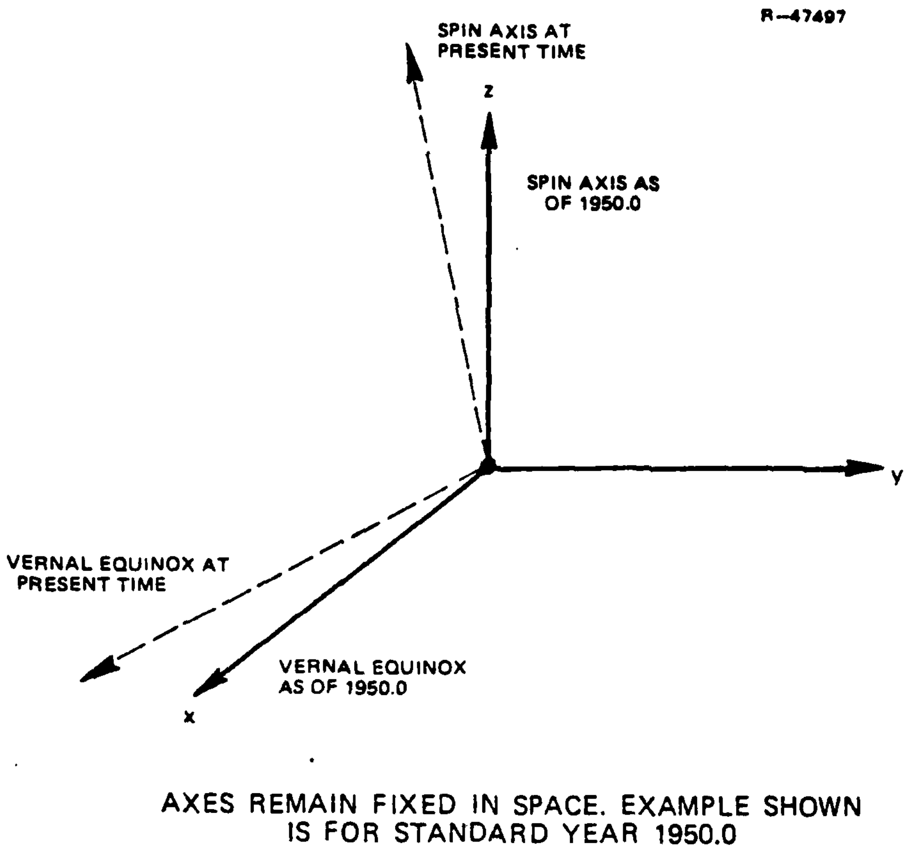


Figure 1.2-21 Coordinate Frames: *True as of a* Specific Date

example, "true of 1950.0," where the notation 1950.0 refers to an instant of time near the beginning of the calendar year.*

A further step in the direction of removing non-*inertial effects is to smooth* the motion of the pole (and the equinox) by averaging out the periodic effects of nutation. The mean equinox, mean pole, and mean equator are affected

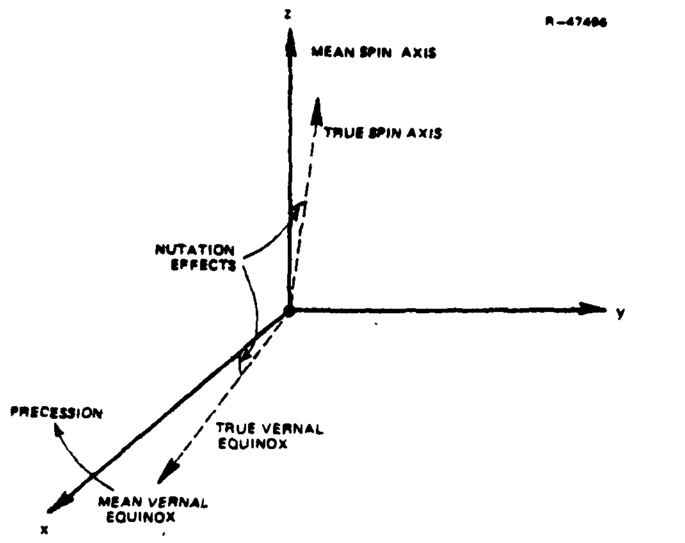
*For the exact definition of notation like 1950.0, the student should refer to the American Ephemeris and Nautical Almanac (AENA), or a textbook on positional (or spherical) astronomy.

only by precession (Fig. 1.2-22). The mean reference frame may be based on the equinox of date (Fig. 1.2-23) or it may be held fixed in space as of a particular instant of time -- e.g., 1950.0. The coordinates relative to these frames are called mean of instant or mean of date in the first case, and mean of 1950.0 -- for example -- in the second case.

Table 1.2-6 summarizes these four kinds of coordinate systems.

The location of any object with respect to one of these coordinate systems is given in terms of polar coordinates analogous to latitude and longitude. The angle indicating distance north or south of the equator is called declination and is measured in degrees, from -90 deg (south) to +90 deg (north). The angle corresponding to longitude (measured positive toward the east) is the right ascension. It is customary to measure this angle in time-like units, with 24 hours corresponding to 360 deg. Right ascension and declination are illustrated in Fig. 1.2-24, while some examples of the correspondence between angular measure and time measure for right ascension are shown in Table 1.2-7.

Star positions, in particular, are listed in star catalogs in terms of their declination and right ascension. Table 1.2-8 reproduces a segment of a star catalog from the American Ephemeris and Nautical Almanac (AENA), with the coordinates shown as mean coordinates for 1980.0 (corresponding to Fig. 1.2-24). It is important for the student to understand that it is actually a collection of star positions -- a star catalog -- that defines the axes of the coordinate system. A considerable effort is expended at observatories around the world on careful measurement and remeasurement of stellar coordinates, in order to refine the definition of the various stellar inertial reference frames.



AXES SHIFT WITH RESPECT TO FIXED DIRECTIONS IN SPACE, BUT SHIFT DOES NOT INCLUDE NUTATION EFFECTS.

Figure 1.2-22 Coordinate Frames: Mean of Instant

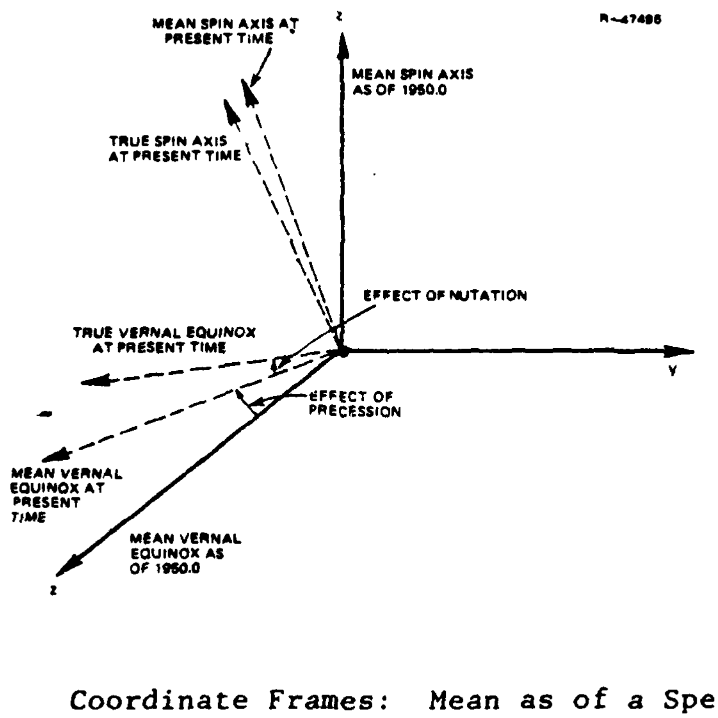


Figure 1.2-23 Coordinate Frames: Mean as of a Specific Date

TABLE 1.2-6

TYPES OF COORDINATE SYSTEM

IDENTIFICATION	AXES MOVE?		AXES FIXED IN SPACE?
	PRECESSION	NUTATION	
True of instant	yes	yes	no
True as of a specific date	no	no	yes
Mean of instant	yes	no	no
Mean as of a specific date	no	no	yes

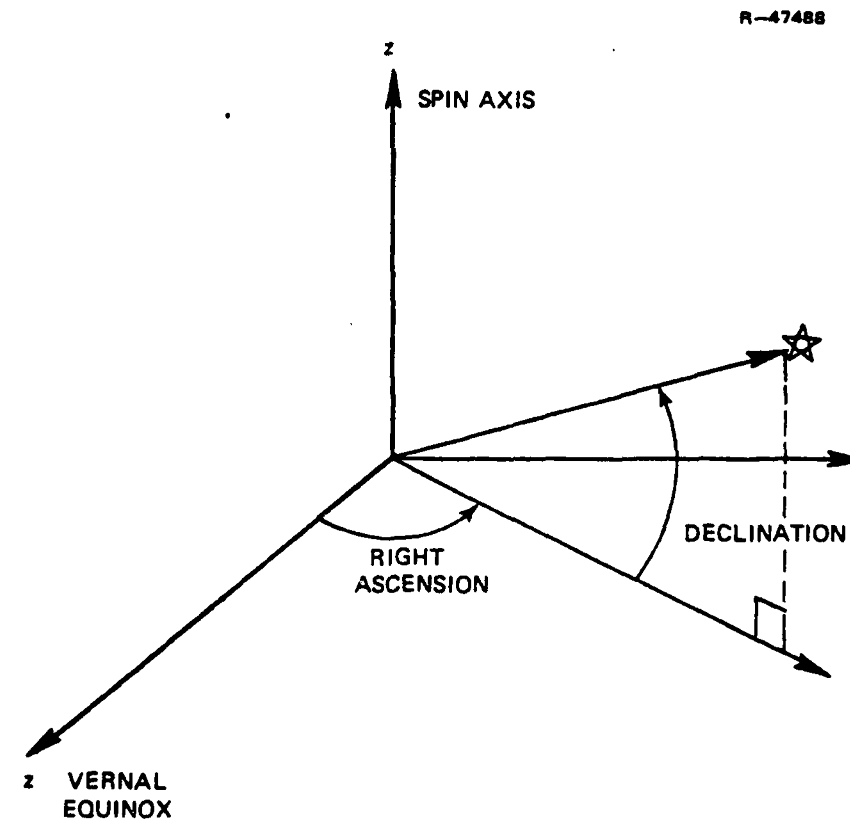


Figure 1.2-24 Right Ascension and Declination

TABLE 1.2-7

RIGHT ASCENSION IN TIME-LIKE UNITS

ANGLE IN DEGREES	ANGLE IN TIME-LIKE UNITS
1 second of arc	$0\overset{s}{.}067$
1 minute of arc	4^s
1 degree	4^m
5 degrees	20^m
20 degrees	1^h20^m
90 degrees	6^h
360 degrees	24^h

<u>Note</u> the convention that <u>hour</u>, <u>minute</u>, and <u>second</u> are indicated as superscript letters.

TABLE 1.2-8

SEGMENT OF A STAR CATALOG

NAME	MAG.	SP.	RIGHT ASCENSION	DECLINATION
			h m s	° ' "
ν Gem	4.1	B5	6 27 46.5	+20 13 33
β Mon	4.6	B2e	6 27 51.2	- 7 01 14
4 CMa	4.3	B1	6 31 01.3	-23 24 11
13 Mon	4.5	A0p	6 31 49.3	+ 7 20 55
ξ^2 CMa	4.5	A0	6 34 13.0	-22 56 54
N Car	4.4	A0	6 34 32.1	-52 57 32
ν CMa	4.1	K0	6 35 48.7	-19 14 17
γ Gem	1.9	A0	6 36 33.4	+16 25 03
8 CMa	4.6	K0	6 37 00.6	-18 13 09
ν Pup	3.2	B8	6 37 08.9	-43 10 40
S Mon	4.7	Oe5	6 39 52.6	+ 9 54 55
ε Gem	3.2	G5	6 42 42.1	+25 09 07
30 Gem	4.6	K0	6 42 51.6	+13 14 57
ξ Gem	3.4	F5	6 [illegible] 10.0	+12 55 05
α CMa	-1.6	A0	6 44 16.0	-16 41 16
18 Mon	4.7	K0	6 46 49.1	+ 2 26 06
α Pic	3.3	A5	6 47 59.2	-61 55 11
κ CMa	3.8	B2p	6 49 05.6	-32 29 05
A Car	4.4	G5	6 49 25.2	-53 35 54
τ Pup	2.8	K0	6 49 25.3	-50 35 26

Notes:

1. Mag. refers to the magnitude, a logarithmic measure of the brightness of the star.
2. Sp. refers to the spectral class, or color, of the star.
3. The segment includes Sirius (α CMa), the brightest star in the sky.

1.2.4.3 Astronomic Positioning

The position of a point on the surface of the earth (latitude and longitude) can be obtained directly by observing the stars. Astronomic positioning is the oldest positioning method. It has been used for many years by mariners and, more recently, by aviators for navigational purposes. Explorers have relied on the astronomic method to locate themselves in uncharted areas. Geodesists use positions determined by astronomic methods along with other types of survey data (such as triangulation and trilateration) to establish precise positions. A pair of isolated astronomic positions not interconnected by geodetic surveys cannot ordinarily be related to one other (for the computation of distance and direction between points) with the accuracy required for modern military applications.

The principle of the astronomic method is to measure the angles relating the local vertical (plumb line) at a survey point to the apparent position of an identified star. Included is the precise time at which the measurements are made. By combining these data with information from a star catalog, which provides the coordinates of the star in a stellar inertial reference frame, the surveyor determines the direction of the plumb line (zenith direction) in the inertial frame. From these data he can then deduce local latitude.

As an illustration, a simple method of latitude determination is presented. In the northern hemisphere, this would consist of measuring the elevation of Polaris (the North Star) above the horizon of the observer. Since Polaris is not exactly aligned with the earth's spin axis, a correction must be applied, as determined (once the time is known) from standard navigational tables. The geometric relations involved in the definition of latitude are shown in Fig. 1.2-25.

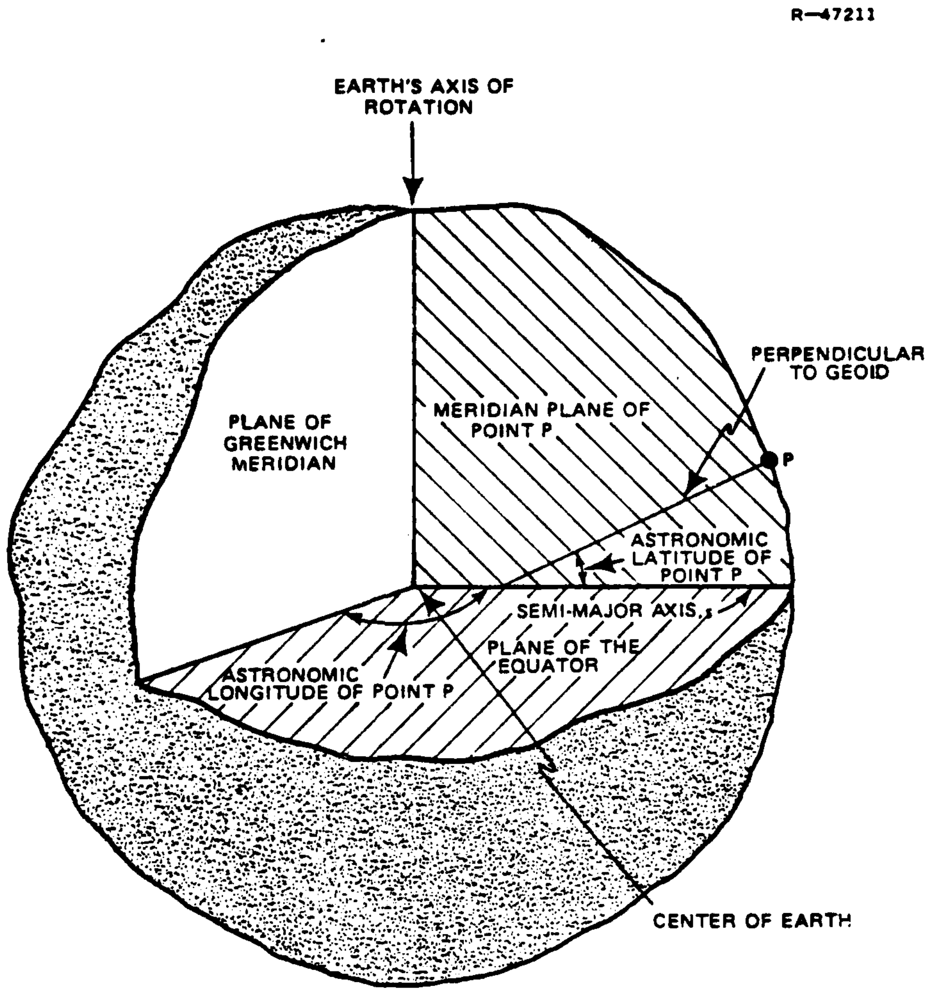


Figure 1.2-25   Astronomic Coordinates

The definition of longitude at a point involves the concept of the local meridian. As shown in Fig. 1.2-25, this is the plane defined at the point by:

- The earth's spin axis
- The local vertical (plumb line, or perpendicular to the geoid).

Longitude is then defined as the angle between the plane of the meridian at Greenwich (Prime Meridian) and the local astronomic meridian of the point. The actual measurement of astronomic longitude is really a time measurement: the time difference between the passage of an identified star over the meridian of Greenwich (as calculated from the star's coordinates and other astronomical reference data) and its observed passage over the local meridian, as timed by an accurate clock calibrated against international standard time transmissions. Time differences are translated into longitude differences on the basis of the earth's rotation rate with respect to the stars (one complete rotation in approximately 23 hours 56 minutes).

Astronomic observations are made by optical instruments -- theodolite, zenith camera, prismatic astrolabe -- which all contain leveling devices. When properly adjusted, the vertical axis of the instrument coincides with the direction of gravity and is, therefore, perpendicular to the geoid. Thus, astronomic positions are referenced to the geoid, and must be distinguished from geodetic positions, referenced to an ellipsoid (Fig. 1.2-26). The astronomic and geodetic coordinates (latitude and longitude) of a point are related to one another by the deflection of the vertical* -- the angular difference between the normal to the geoid and the normal to the ellipsoid. The difference in direction between these two normals is usually described in terms of its north-south component, ξ, and its east-west component, η. Astronomic and geodetic coordinates are then related by the equations

*This concept has been introduced in Section 1.2.1.

R—47210

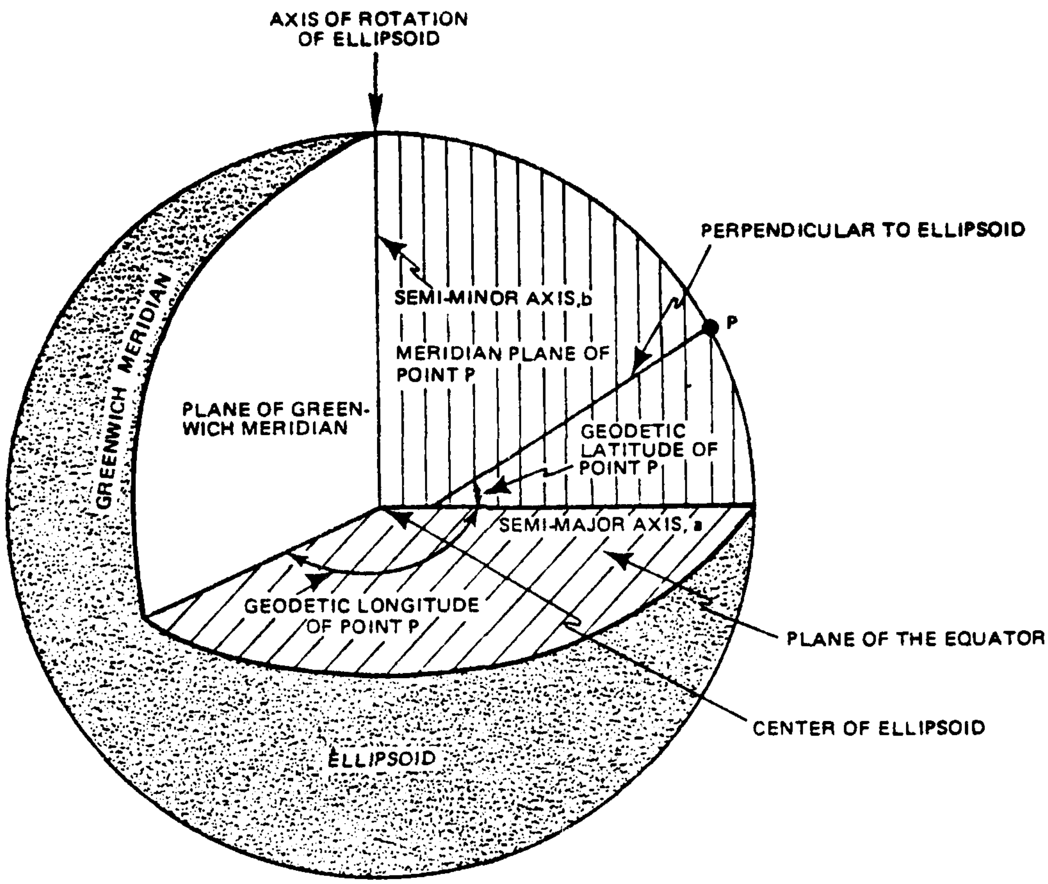


Figure 1.2-26   Geodetic Coordinates

$$\xi = \Phi - \phi \qquad (1.2\text{-}5)$$

$$\eta = (\Lambda - \lambda) \cos \phi \qquad (1.2\text{-}6)$$

where

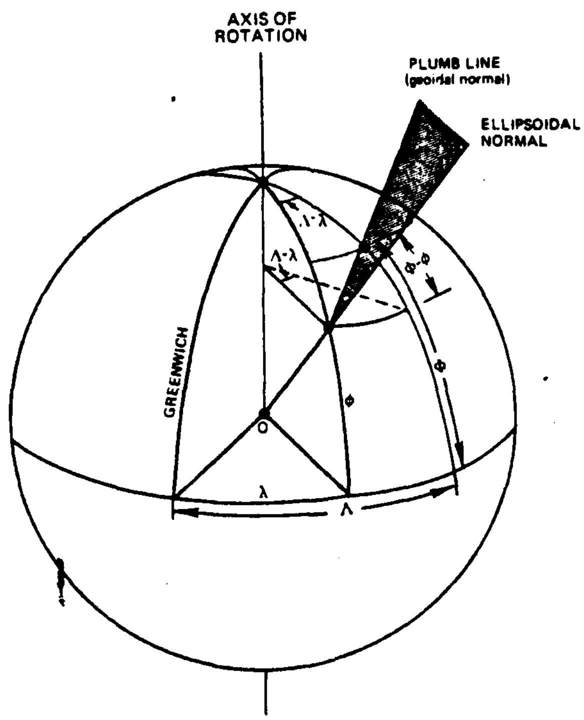
Φ = astronomic latitude

ϕ = geodetic latitude

Λ = astronomic longitude

λ = geodetic longitude

These relationships are shown in Fig. 1.2-27. The student should bear in mind that deflections of the vertical are measured in seconds of arc, and -- in the <u>worst cases</u> -- can exceed one minute of arc. But it is essential to take them into account for accurate positioning. Recall from Section 1.2.1 that the geoid is an irregular surface, not defined in tractable mathematical form; hence, on the basis of the <u>astronomic</u> coordinates of two points of interest, it is not feasible to describe their <u>exact</u> geometric relationship (for example, direction and distance). Geodetic coordinates, on the other hand, refer to an exactly defined ellipsoid. If the geodetic coordinates are known, the spatial relationship between two points (launch point and target, for example) is defined exactly by the geometry of the ellipsoidal surface.



Λ = astronomic longitude
λ = geodetic longitude
Φ = astronomic latitude
ϕ = geodetic latitude

ξ = north-south component of deflection of the vertical = $\Phi - \phi$
η = east-west component of deflection of the vertical = $(\Lambda - \lambda) \cos \phi$
O = center of the earth

Figure 1.2-27 Relation Between Astronomic and Geodetic Coordinates

It should be emphasized that for most applications requiring high precision, astronomic coordinates are not usable unless they can be reduced to the corresponding geodetic coordinates by applying corrections:

- For the deflection of the vertical to latitude and longitude
- For the undulation (radial separation between ellipsoid and geoid) to the height above sea level

in order to obtain geodetic latitude and longitude, as well as height above the ellipsoid. Thus, local gravity-field information is required in order to make geodetic use of astronomical results, pointing out -- once again -- the intimate connection between the fields of geometric geodesy (size, shape, and relative locations) and physical geodesy (structure of the gravity field).

This brief review of astronomic positioning has emphasized principles at the expense of details. As one example, the astronomically observed latitude and longitude refer to the instantaneous pole (direction of the earth's spin axis). Since the spin axis is not fixed with respect to the earth, but undergoes small, roughly periodic motions around a mean position with a principal period somewhat greater than a year (see Section 1.2.4.1), it is necessary to reduce observed latitude and longitude to a standard mean pole, defined by international agreement, using tabulated polar motion data.

1.2.5 Datums

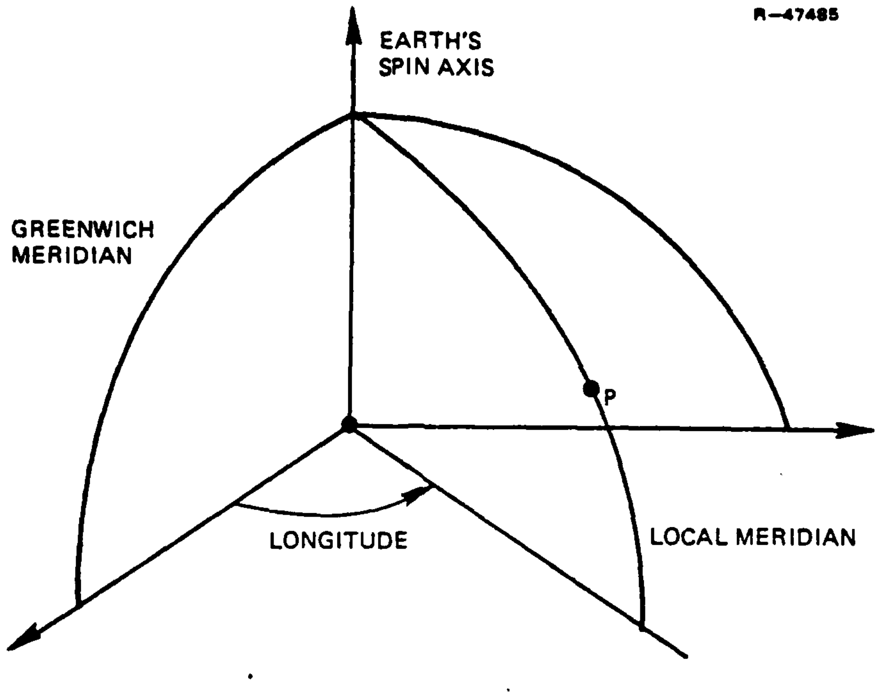
A datum, in geodesy, is a coordinate system for expressing the locations of points on (or near) the surface of the earth. Since (as discussed in Section 1.2.1) the earth's

surface is closely approximated by an ellipsoid, it is generally taken for granted that the system will involve geodetic latitude and longitude based on a particular ellipsoid. These terms, defined in Section 1.2.1, are reviewed in Fig. 1.2-28. The datum, therefore, is specified by the choice of an ellipsoid and the association of that ellipsoid with one or more points on the earth's surface.

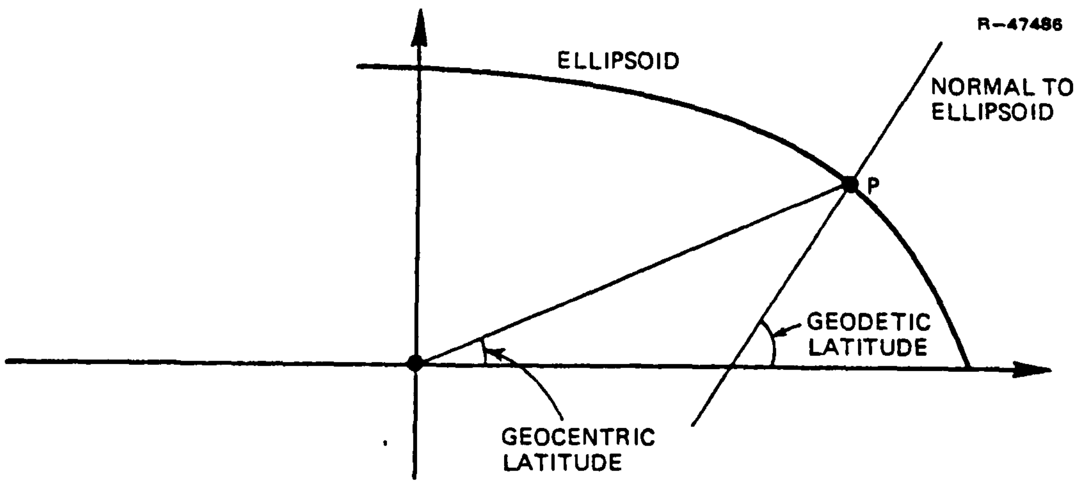
From the historical point of view, datums were originally selected to express locations only within a particular region. Such regional datums were based on an ellipsoid chosen to fit the geoid within that particular region, and oriented with respect to a specific point (origin) within the region. More recently, interest has shifted to the development of global datums, with an ellipsoid chosen to fit the geoid in a global sense, and having its center at the earth's center of mass.

In addition to the distinction between regional and global datums, it is also useful to distinguish between horizontal and vertical datums. The horizontal datum provides a basis to which horizontal coordinates (latitude and longitude) of a position are related; the vertical datum provides a basis to which elevations are related. Most of the discussion in this section pertains to the subject of horizontal datums, with a brief reference to vertical datum control concluding the section.

How a datum is defined - The specification of a particular datum involves an ellipsoid, an origin, and a way of aligning the ellipsoid with the geoid at the origin. More specifically, the definition involves:



(a) GEODETIC LONGITUDE AT POINT P



(b) GEODETIC LATITUDE AT POINT P

Figure 1.2-28 Geodetic Latitude and Longitude

- <u>Ellipsoid</u> -- determined by its semimajor axis, a, and flattening, f

- <u>Origin</u> -- a specific point, centrally located, which will serve as the base station for a survey network

- <u>Alignment</u> -- at the origin, the relative orientations of ellipsoid and geoid are defined by

  - the relative deflection between the normal to the ellipsoid and the normal to the geoid (plumb line), usually expressed in terms of its North-South and East-West components. These may be taken to be zero, in which case geoid and ellipsoid are locally parallel.

  - the geoid height or undulation at the origin (separation between geoid and ellipsoid, measured along the normal to the ellipsoid). The undulation is also sometimes set to zero, making geoid and ellipsoid locally tangent.

  - the azimuth from the origin to another specified point, establishing the first line of a survey network.

The final condition usually imposed is that the minor axis of the ellipsoid is parallel to the earth's axis of rotation.

Some examples of <u>regional datums</u> are given in Table 1.2-9; a few <u>global datums</u> are listed in Table 1.2-10. Four of the regional datums will be discussed in more detail below.

TABLE 1.2-9

EXAMPLES OF REGIONAL DATUMS

T-3678

DATUM NAME	ORIGIN	ELLIPSOID NAME	ELLIPSOID PARAMETERS	
			SEMIMAJOR AXIS (m)	FLATTENING
North American Datum - 1927 (NAD 27)	Meades Ranch, Kansas	Clarke (1866)	6378206	1/295.0
European Datum (ED)	Potsdam (near Berlin), Germany	International (Hayford)	6378388	1/297.0
Russian Datum (Pulkovo 42)	Pulkova Observatory Leningrad, USSR	Krasovskii (1938)	6378245	1/298.3
Tokyo Datum (TD)	Tokyo, Japan	Bessel (1841)	6377397	1/299.2
Indian Datum	Kalianpur, India	Everest	6377276	1/300.8

TABLE 1.2-10

EXAMPLES OF GLOBAL DATUMS

T-3679

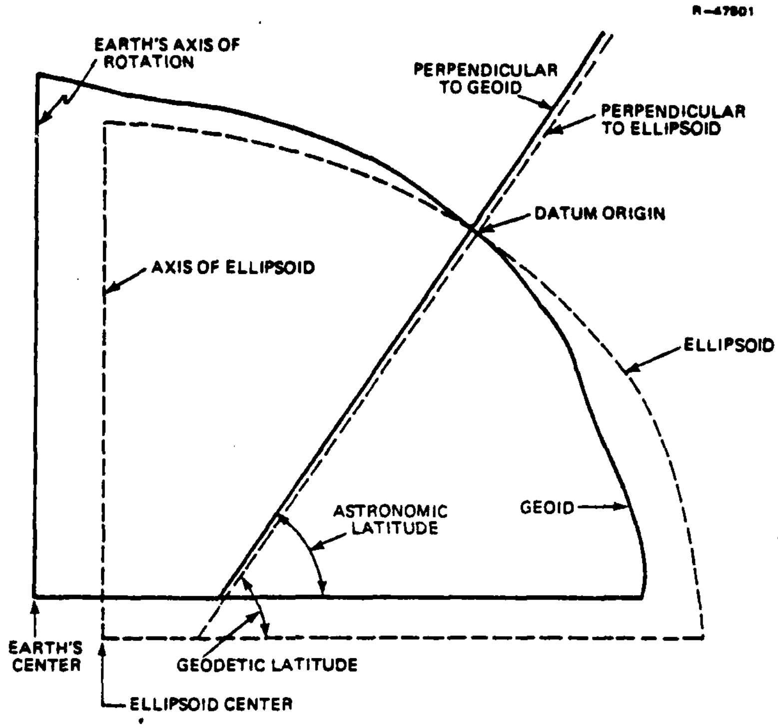
DATUM NAME	ELLIPSOID NAME	ELLIPSOID PARAMETERS	
		SEMIMAJOR AXIS (m)	FLATTENING
Kaula 1961	Kaula	6378165	1/298.30
Modified Mercury Datum 1968 (MMD 68)	Fischer (1968)	6378150	1/298.30
Smithsonian Datum 1966 (SAO 66-C6)	SAO 66	6378165	1/298.00
World Geodetic System 1972 (WGS 72)	WGS 72	6378135	1/298.26

NOTE: No origin is given, since global datums are based on the center of mass of the earth.

How the ellipsoid is oriented to the geoid - Two examples will be given to illustrate the basic principles of geoid - ellipsoid alignment. In the first method, the alignment is based entirely on the point selected as origin. The second method involves a best-fit solution over a number of points.

It is helpful to recall the definitions of a few terms introduced in Section 1.2.4.3 (Astronomic Positioning). When latitude and longitude are determined by star sightings, using the local horizon and local vertical (plumb line, normal to the geoid), together with the earth's axis of rotation, as a directional reference, they are called astronomic latitude and longitude. Azimuths measured using the astronomically defined north and vertical directions are likewise called astronomic azimuths.

Single-position datum orientation is based on the selection of a station, usually one located near the center of a triangulation network, to serve as the datum origin. Then the astronomical coordinates of the station and the astronomical azimuth of a line from the station to another control station are observed. The observed astronomical coordinates and azimuth are adopted, without any correction, as the geodetic coordinates and azimuth of the datum origin on the reference ellipsoid. Usually, the geoid and ellipsoid are assumed to coincide at that point. This means that the deflection of the vertical and the separation (or undulation) between the ellipsoid and geoid are defined as zero at the origin. By this choice of orientation, the normal to the ellipsoid is arbitrarily made to coincide with the plumb line at the datum origin. Figure 1.2-29 illustrates the relation between geoid and ellipsoid when the orientation is based on a single station.



Note: Perpendicular to ellipsoid coincides with perpendicular to geoid at datum origin.

Figure 1.2-29 Datum Orientation Based on Single Station

Although the computed positions are correct with respect to each other in this type of orientation, the entire net is shifted with respect to the axis of the earth. This is not significant for local use of the positions but may introduce large systematic errors far from the origin.

Note that although the deflections of the vertical and undulation are defined as zero at the origin, deflections occur at other positions within the network. When the geodetic latitude and longitude of other points in the net are compared with the corresponding astronomical latitude and longitude, differences appear between the two sets of values.

Also, positions derived from different astronomically oriented datums are not directly comparable to each other in any geodetic computation. The Tokyo Datum is an example of this type of datum orientation.

Multiple-Station datum orientation is based on the best average fit between the ellipsoid and the geoid at several selected survey stations in the area for which the datum is designed. The principles are illustrated in Fig. 1.2-30.

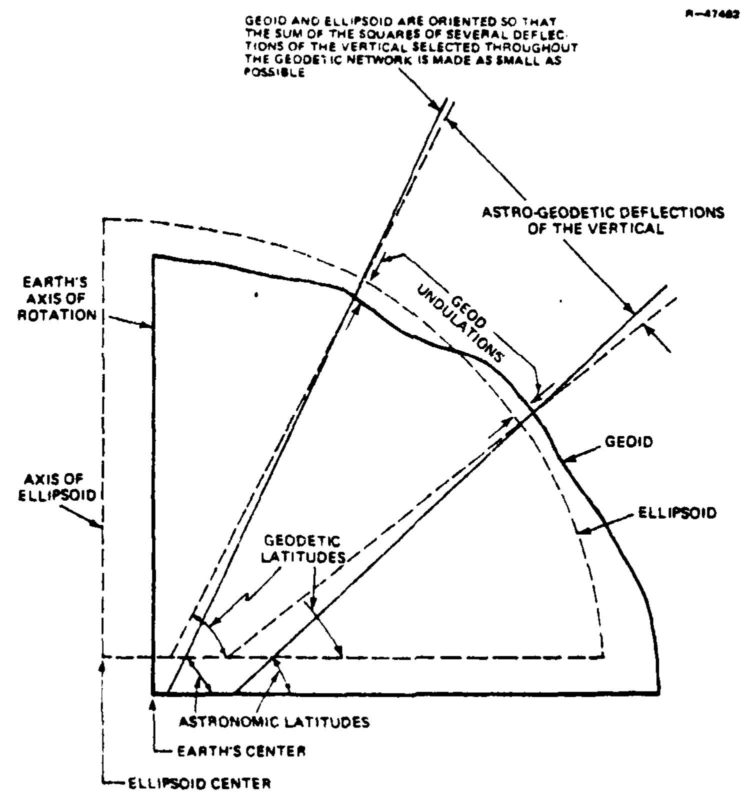


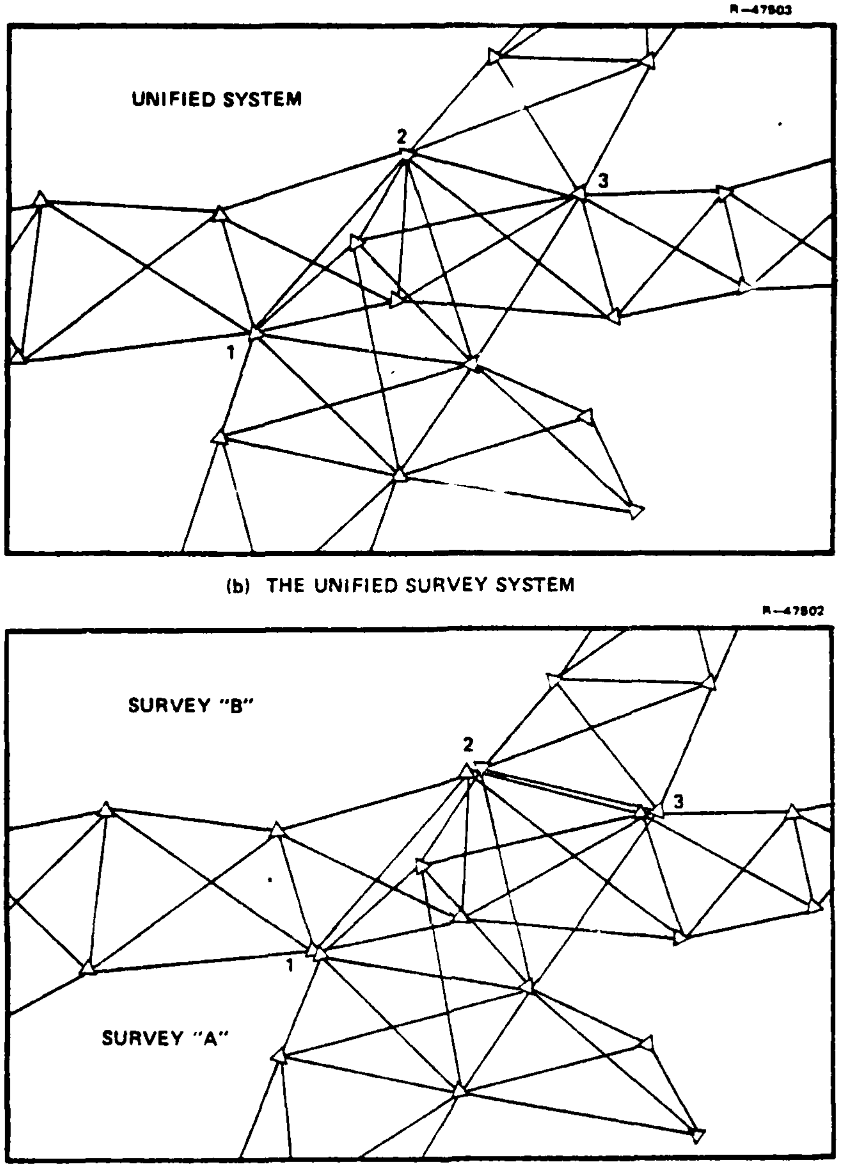
Figure 1.2-30 Datum Orientation Based on Multiple Stations

Discrepancies between datums - In areas of overlap between two geodetic survey networks, each computed with respect to a different datum, the coordinates of any particular point given with respect to one datum differ from those given

with respect to the other. The differences occur because of the different ellipsoids used. There is possible offset between the centers of the ellipsoids, relative rotation between the systems, and differences of scale. As a consequence, the computation of geodetic information from one datum to another unconnected datum is virtually impossible, regardless of the accuracy of the individual datums for internal computations.

With the development of both intermediate and long-range weapon systems, these geodetic problems have taken on major military significance. To satisfy military requirements, it is necessary to provide detailed cartographic coverage of areas of strategic importance and to perform geodetic computations between these areas and launch sites which are often on unrelated datums. Both of these requirements call for the unification of major datums; this is done by the use of one (or a combination) of several existing methods.

Datum connection - There are three general methods by which horizontal datums can be connected. The first method is restricted to surveys of a limited scope and consists of systematic elimination of discrepancies between adjoining or overlapping geodetic survey networks. This is done by moving the origin, rotating, and stretching networks to fit each other. The method, known as datum transformation or datum reduction, is used to connect local surveys for mapping purposes. Figure 1.2-31 illustrates the basic principles. The second method, called the celestial method, uses observations of astronomical bodies or events (eclipses, occultations) from points within the two networks to effect the connection between them. These methods are now largely superseded by techniques of satellite geodesy (described in more detail in Chapter Four). The third approach involves the use of gravimetric data to tie the individual datums into a unified world system. This will be discussed later (Unit Two) under the heading World Geodetic Systems.



(a) ORIGINAL SURVEYS

Figure 1.2-31 Datum Reduction Applied to Overlapping Survey Systems

<u>Major local datums before World War II</u> - By 1940, every technically advanced nation had developed its own geodetic system to an extent governed by its economic and military requirements. Some systems were developed by the expansion and unification of existing local surveys and others by new nationwide surveys replacing outdated local ones. Normally, neighboring countries did not use the same geodetic

datum. Not only were there no economic requirements for common geodetic information, but the use of common datums was often seen as contrary to the military interests of the individual countries. The only surveys of an international nature based on one datum were the few measurements of long arcs accomplished for the purpose of determining the size and shape of the earth. The net result was that there were many different surveys of varying size which differed from each other remarkably. The national maps based on the surveys also differed widely. Figure 1.2-32 illustrates the profusion of geodetic datums in Southeast Asia in 1940.

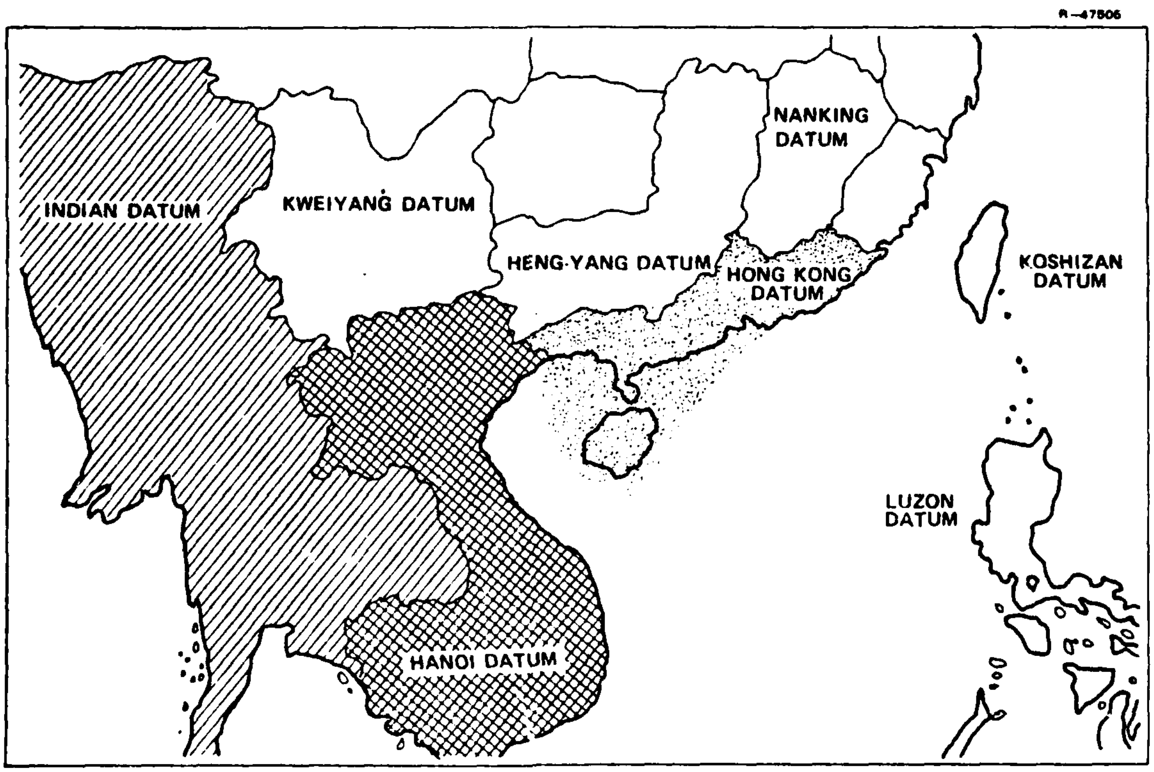


Figure 1.2-32 Example of Different Geodetic Datums in Southeast Asia

As military distance requirements increased, positioning information of local or even national scope became unsatisfactory. The capabilities of the various weapon systems increased until datums of at least continental limits were required. The solution that evolved was the establishment of a <u>preferred datum</u> for an area and the adjustment of all local systems to it. The North American, European, Tokyo, and Indian Datums were (initially) selected for this purpose. Figure 1.2-33 illustrates, in a general manner, the coverage of these four datums. Other major geodetic datums of the world include the Arc and Adindan Datums in Africa, the Australian Geodetic Datum, South American 1969 Datum, and the Russian Pulkovo 1942 Datum. The four preferred datums are now reviewed.

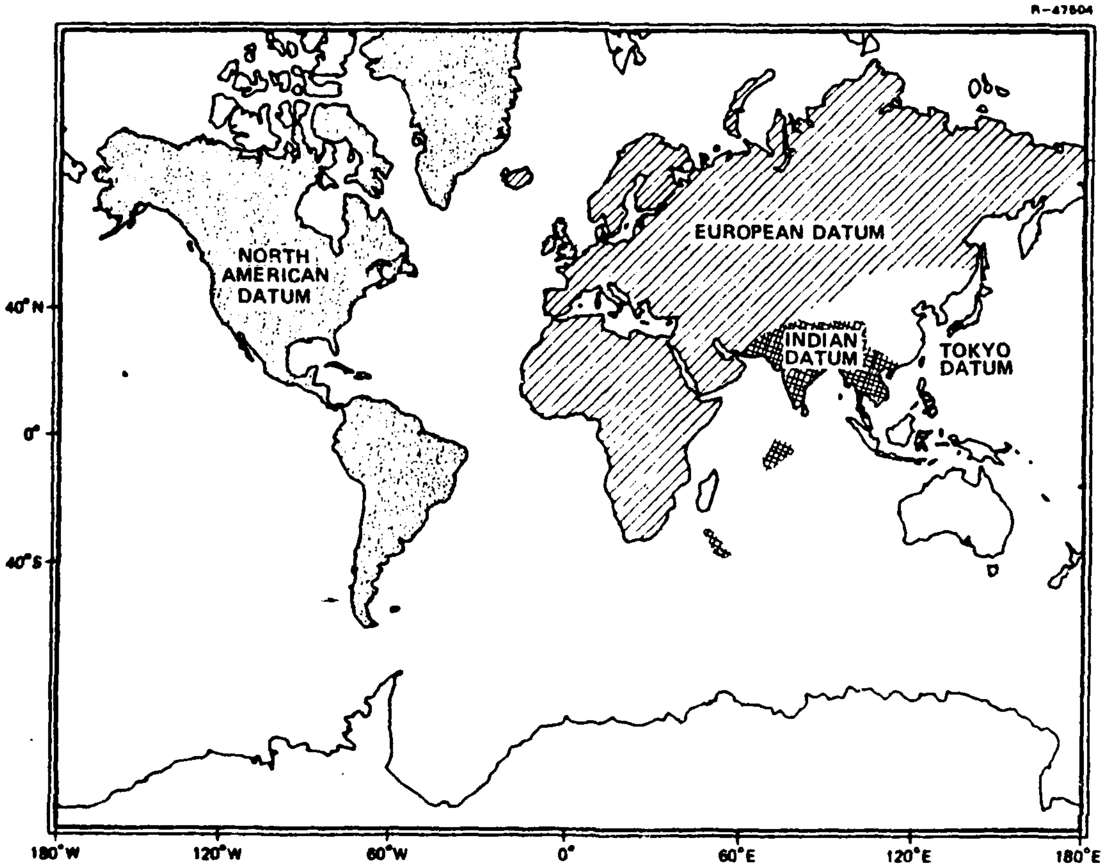


Figure 1.2-33 The Preferred Datums

The preferred datums - The North American Datum (1927) has evolved from the first official geodetic datum in the United States, the New England Datum, adopted in 1879. It was based on surveys in the eastern and northeastern states and referenced to the Clarke 1866 Ellipsoid. Through the years this datum was extended to the south and west and, in 1901, the expanded network was officially designated the United States Standard Datum. The survey station at Meades Ranch in Kansas was selected as the origin, with adopted geodetic coordinates:

$$\text{Latitude} = 39° \; 13' \; 26.''686 \text{ N} \tag{1.2-7}$$

$$\text{Longitude} = 98° \; 32' \; 30.''506 \text{ W} \tag{1.2-8}$$

The initial azimuth was between Meades Ranch and the survey station designated Waldo, with a value of

$$\text{Azimuth} = 75° \; 28' \; 09.''64 \tag{1.2-9}$$

In 1913, Canada and Mexico formally agreed to base their survey networks on the United States system. The datum was then renamed the North American Datum. Adjusting new surveys to fit into the network created many problems and, therefore, during the five-year period 1927-1932 all available first-order data were adjusted into a system now known as the North American 1927 Datum. The datum is computed on the Clarke 1866 Ellipsoid, which was oriented by a multiple-station method. The system not only incorporates Canada and Mexico but has connections to the South American Datum 1969 through the West Indies and Central America.

The European Datum has its origin at Potsdam, Germany (just outside of Berlin). Numerous European national systems were joined into a large datum based upon the International

Ellipsoid, oriented by multiple-station methods. The U.S. Army Map Service, now known as the Defense Mapping Agency Hydrographic Topographic Center (DMAHTC), connected the European and African geodetic survey chains and filled the gap in the African arc measurement from Cairo to Cape Town. This work related the Adindan Datum in North Africa, which roughly follows the Twelfth Parallel, and the Arc Datum, extending from the Equator to the African Cape, to the European Datum. Through common survey stations, a datum transformation was derived between the old Russian Pulkovo 1932 and European systems. This extended the European Datum eastward to the 84th meridian. In 1946 the Pulkovo 1932 system was united with a basic Siberian network and the new datum was designated the <u>1942 Pulkovo System of Survey Coordinates</u> (Pulkovo Datum 1942). Additional ties across the Middle East connected the European with the Indian Datum.

<u>The Indian Datum</u> has been accepted as the preferred datum for India and several adjacent countries in Southeast Asia. It is computed on the Everest Ellipsoid with its origin at Kalianpur in Central India. Derived in 1830, the Everest Ellipsoid is the oldest of the ellipsoids in use and is much too small from a global point of view. As a result, the datum cannot be extended too far from the origin because very large geoid separations occur. For this reason and the fact that the ties between local surveys in Southeast Asia are typically weak, the Indian Datum is probably the least satisfactory of the preferred datums.

<u>The Tokyo Datum</u> is the fourth of the initially selected preferred datums. It is defined in terms of the Bessel Ellipsoid and oriented by means of a single astronomic station. With survey ties through Korea, the Tokyo Datum is connected with the Manchurian Datum. Unfortunately, Tokyo is situated

on a steep geoid slope and the single-station orientation has resulted in large systematic geoid separations as the system is extended from its initial point.

For military distance and direction problems limited to continental areas (or smaller), the preferred datums are satisfactory. However, while they are improvements over the limited national datums, they do not provide the precise geodetic information required for intercontinental ballistic missiles.

As an example, the European and North American Datums have been connected by electronic surveying techniques (the North Atlantic HIRAN tie), but the required level of precision is still not attained. For each of these datums, the ellipsoid chosen is an adequate fit in the area of the origin, but neither ellipsoid provides a good fit for the entire earth. Also, the process of connecting various datums by means of intervening datums or survey ties allows errors to accumulate which do not always provide agreement with newly observed data. The surveys joining India to the European and the Tokyo Datums present similar problems. Further discussion of this problem is deferred to Unit Two (World Geodetic Systems).

Vertical datums - Just as horizontal surveys are referred to specific original conditions (datums), vertical surveys are also related to an initial quantity or datum. It is customary to refer elevations to the geoid (rather than the ellipsoid) because the instruments used either for differential or trigonometric leveling (see Section 1.2.2, Vertical Control) are adjusted with the vertical axis coincident with the local vertical. As with horizontal datums, there are many discrepancies among vertical datums. However, the root mean

square (RMS) difference between leveling nets based on different mean sea level datums can be as large as two meters. Elevations in some areas are related to surfaces other than the geoid; errors are larger in such areas.

In the European area, there are fewer vertical datum problems than in Asia and Africa. Extensive leveling work has been done in Europe and practically all of it has been referred to the same mean sea level surface. However, in Asia and Africa the situation has been different. There is very little reliable, recent, vertical data available for much of the area of these continents. In places there is precise leveling information available based on mean sea level. In other areas the zero elevation is an assumed elevation which sometimes has no connection to any sea level surface. China is an extreme example of this situation, since nearly all of the provinces have an independent zero reference.

The mean sea level surface for the United States was determined using 21 tidal stations in this country and five in Canada. This vertical datum has been extended over most of the continent by first-order differential leveling.

1.2.6 Cartographic Applications

The only way to model the relative size, shape, and orientation of features on the surface of the earth -- with complete accuracy, free of any distortion -- is by the use of a globe. For obvious reasons of convenience and portability, it is necessary to depict either the whole earth or specific portions on a flat surface. In this process of map construction, distortions arise inevitably (technically expressed, an ellipsoid is not developable to a plane). The study of these

distortions, and the choice of mapping systems to minimize them, is the subject of map projections.

The position of a point in a planar map is described in terms of rectangular coordinates, as shown in Fig. 1.2-34. Traditionally, the Y-coordinate is called the northing; the X-coordinate, the easting.* The location of a point on the ellipsoid (representing the surface of the earth) is given by the geodetic latitude and longitude (Fig. 1.2-34). A particular projection is defined by the pair of equations

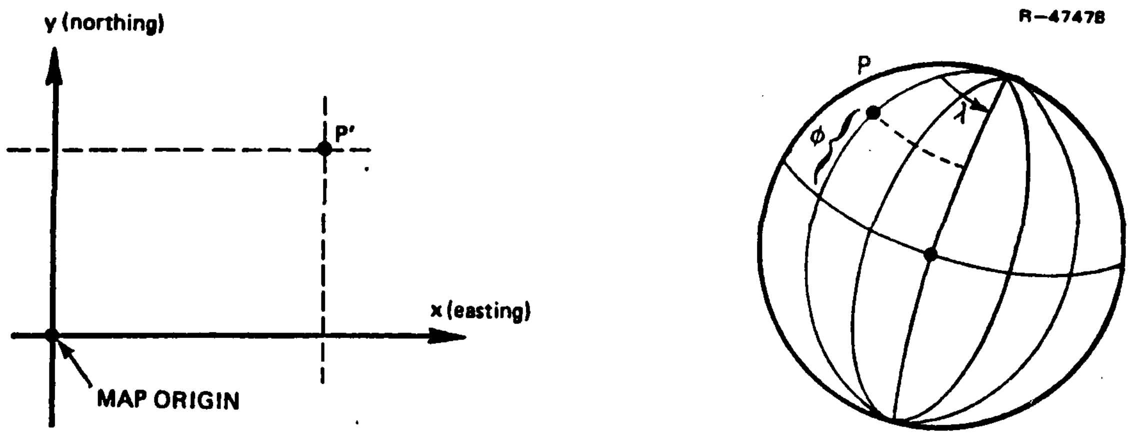


Figure 1.2-34 Map Projection Coordinates

$$X = f(\phi, \lambda) \tag{1.2-10}$$

$$Y = g(\phi, \lambda) \tag{1.2-11}$$

---

*The reader is cautioned that many textbooks on cartography and map projections reverse these meanings of X and Y.

where

- ϕ is the geodetic latitude of a point on the ellipsoid

- λ is the geodetic longitude of a point on the ellipsoid (the same as the geocentric longitude)

- X,Y are the rectangular coordinates of the corresponding point on the map

The simplest possible map projection corresponds to the equations of transformation

$$Y = k\phi \tag{1.2-12}$$

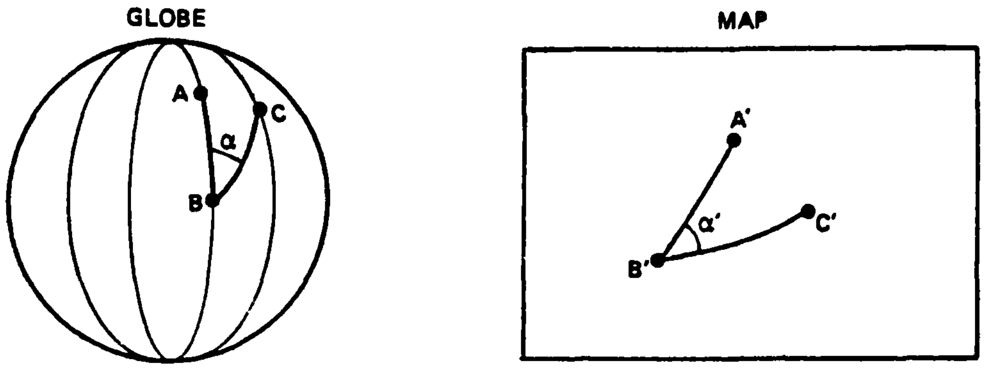
$$X = k\lambda \tag{1.2-13}$$

in which the latitude and longitude are used directly as rectangular coordinates. While this projection, known by its French name, plate carrée, has the advantage of simplicity (and is sometimes used for computer-generated maps), it has three major disadvantages:

- It does not represent distances between points correctly (it is not equidistant)
- It does not represent shapes and angles correctly (it is not conformal)
- It does not represent areas correctly (it is not equivalent).

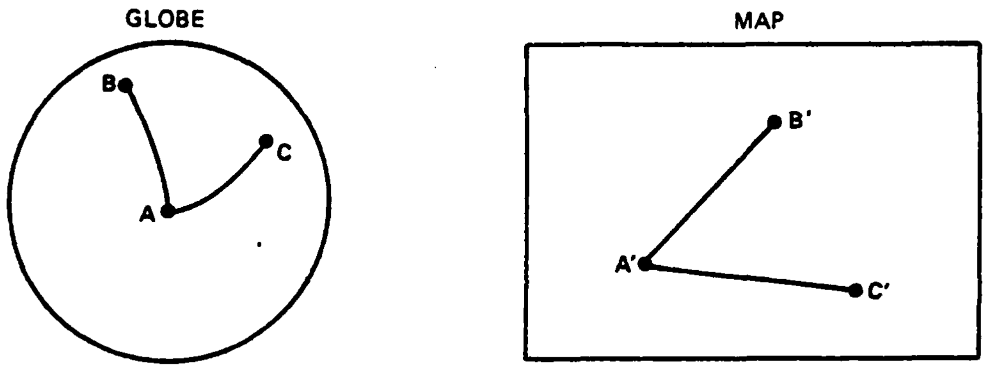
These three forms of distortion are illustrated in Fig. 1.2-35.

R—47651



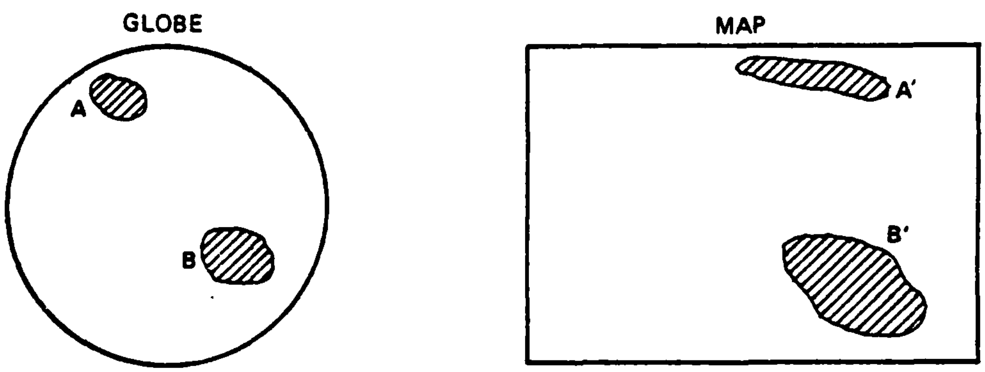
DOES α = α'?
Preservation of angles implies preservation of shapes

a) CONFORMALITY



If AB = AC, are the distances A' to B' and A' to C' equal?

b) EQUIDISTANCE



If areas A and B are equal on the globe, are the areas A' and B' equal on the map?

c) EQUIVALENCE

Figure 1.2-35   Map Distortions

Many important map projections are designed to reduce <u>one</u> of the sources of distortion to zero. These projections can be categorized as

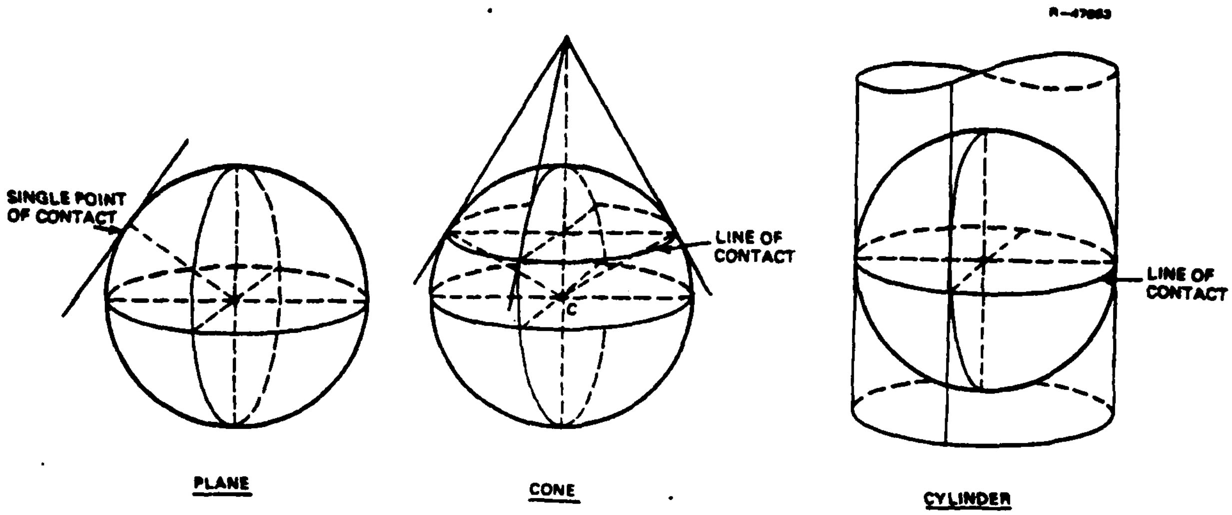
- Equidistant
- Conformal
- Equivalent.

These classifications will be used to organize the map projections to be discussed in this section. For applications in geodesy and navigation, as well as for most other military purposes, <u>conformal</u> maps are most widely used, with <u>equidistant</u> maps also of significance. <u>Equivalent</u> maps are most widely used in geography, and for the depiction of various kinds of statistical and economic data.

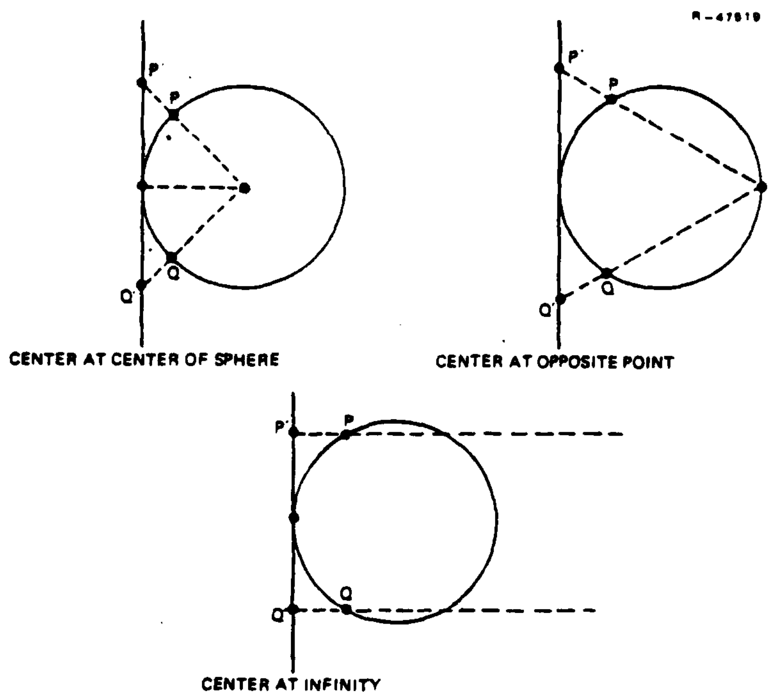
Many (but not all) map projections can also be defined geometrically, in terms of an actual perspective projection onto a <u>plane</u>, cr onto a <u>cone</u> or <u>cylinder</u> (either of which can then be unrolled, or <u>developed</u>, to form a planar map). Projections are then classified according to

- The type of surface projected upon -- plane, cylinder, or cone (see Fig. 1.2-36)
- The location of the center of perspective (see Fig. 1.2-36).

Methods of geometric map construction, based on the characterization of maps according to geometric perspective properties, were formerly of considerable importance when mapping was done entirely by hand. Map reconstruction is now carried out by computer implementation of equations of the form of Eqs. (1.2-10) and (1.2-11), with the output used to control automatic plotting



a) PROJECTION SURFACES



b) DIFFERENT CENTERS OF PERSPECTIVE

Figure 1.2-36 Basic Projection Surfaces and Centers of Perspective

equipment. As a consequence, the geometric interpretation of map projections is of significance primarily for descriptive purposes.

An obvious way to summarize the general nature of any particular map projection is to plot the network of lines (or curves) corresponding to equally spaced meridians of longitude and parallels of latitude. This network is called a graticule. For the plate carrée, for example, the graticule is a network of squares (Fig. 1.2-37).

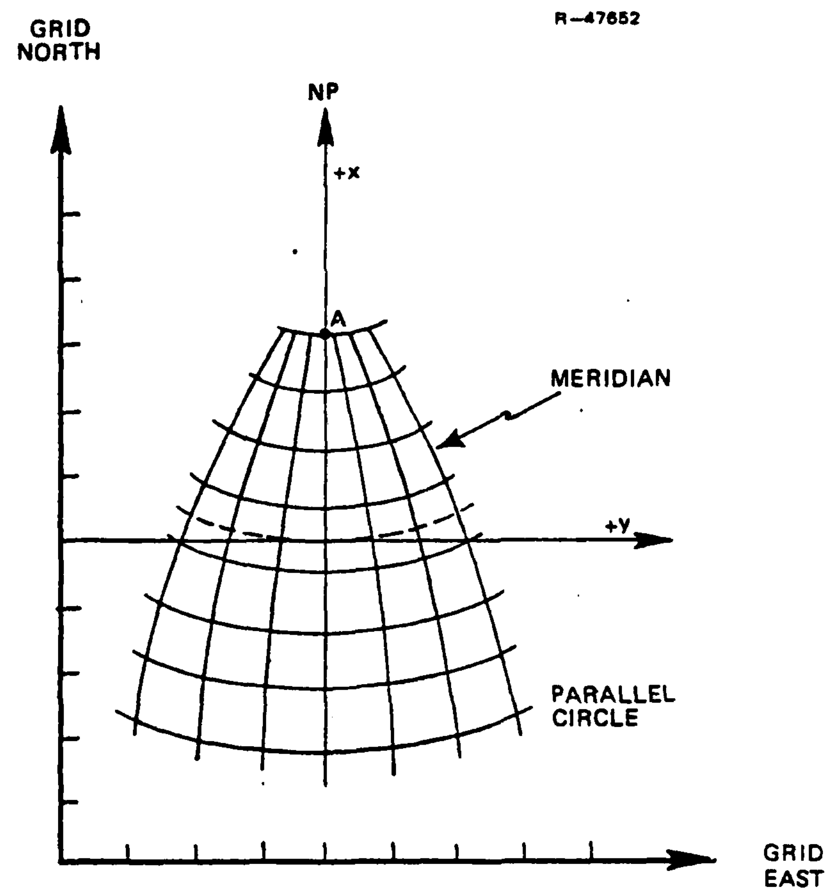


Figure 1.2-37 Example of Graticule

Conformal Projections - Conformal projections are the most widely used maps in navigation and geodesy, because they represent shapes and angles correctly (review Fig. 1.2-35). Three kinds of conformal projections are studied -- those based on perspective projection onto a

- Cylinder (Mercator, Transverse Mercator, and Universal Transverse Mercator)
- Cone (Lambert Conical)
- Plane (Polar Stereographic).

Mercator Projection - The Mercator projection, certainly the best known of all maps, has been widely used for navigation (and other purposes as well) for over 400 years. It may be remembered as the result of two distinct stages:

- Projecting the ellipsoid onto a cylinder (Fig. 1.2-38)
- Stretching the resulting map (in the North-South direction) by an amount that increases with increasing latitude, in order to achieve conformality.

The Mercator projection is defined formally by the equations

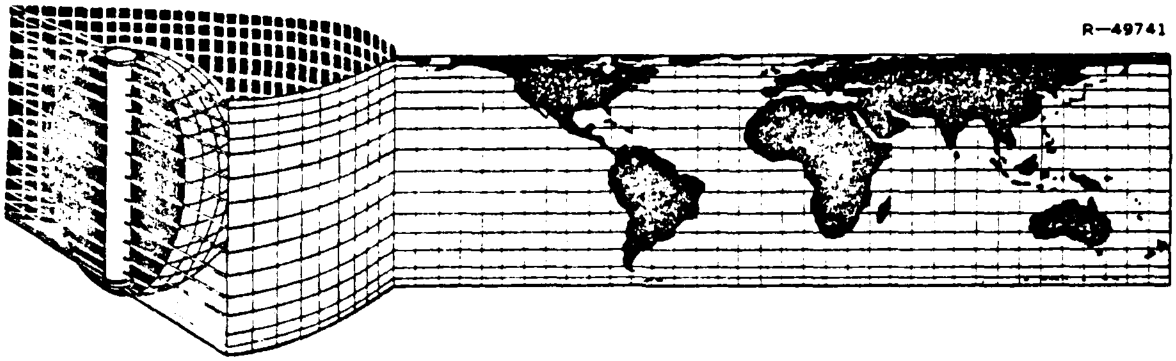
$$Y = k \ln \tan \left(45° + \frac{\phi}{2}\right) \left[\frac{1-e \sin \phi}{1+e \sin \phi}\right]^{\frac{e}{2}} \qquad (1.2\text{-}14)$$

$$X = k\lambda \qquad (1.2\text{-}15)$$

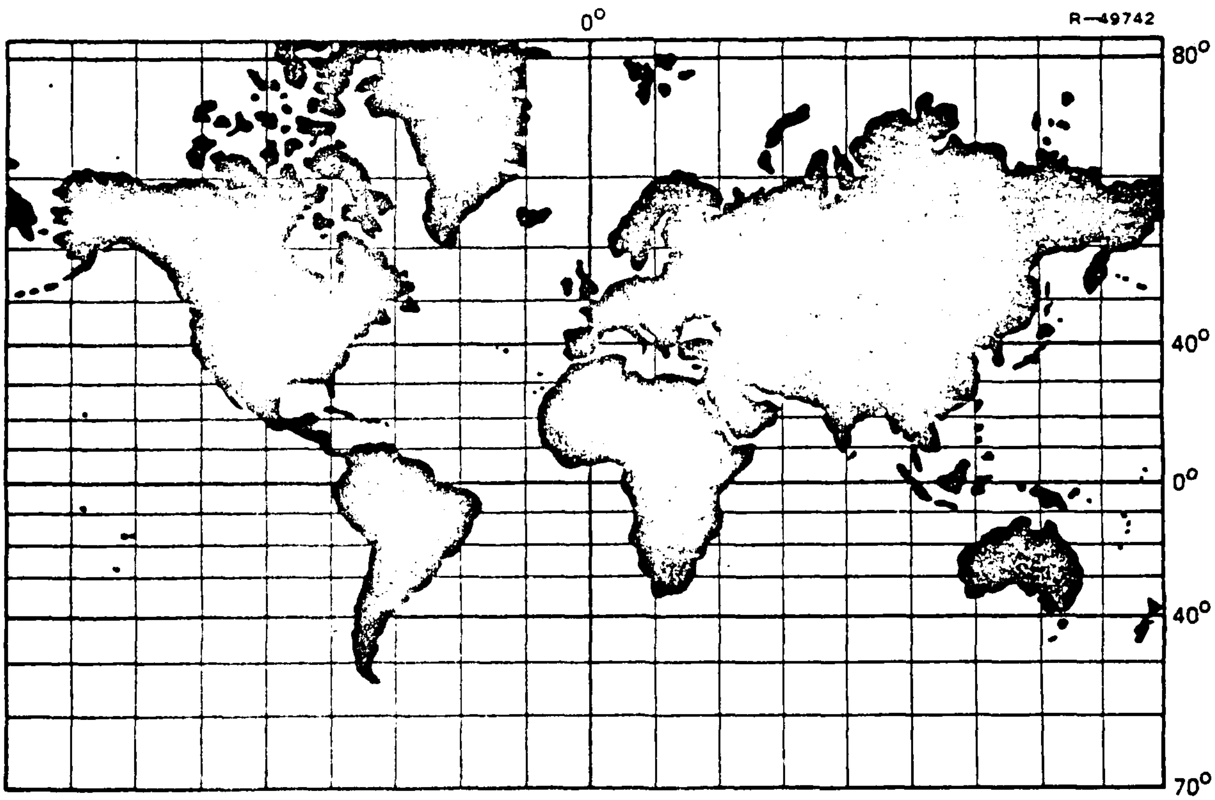
where

X and Y are the rectangular coordinates (northing and easting) of the map

ϕ is the geodetic latitude



a) GLOBE PROJECTED ON CYLINDER



b) STRETCHING TO ACHIEVE CONFORMALITY

Figure 1.2-38   Mercator Projection

λ is the geodetic longitude

e is the eccentricity of the ellipsoid

ln indicates the natural logarithm

The eccentricity of the ellipsoid, e, is related to the flattening, f, by the formula

$$e^2 = 1-(1-f)^2 = f(2-f) \qquad (1.2\text{-}16)$$

and has the approximate value

$$e = 0.082 \qquad (1.2\text{-}17)$$

The bracketed quantity in Eq. (1.2-14) can be expressed, to a good degree of approximation, as

$$\left[\frac{1-e \sin \phi}{1+e \sin \phi}\right]^{\frac{e}{2}} \simeq 1-e^2 \sin \phi \qquad (1.2\text{-}18)$$

and ranges from 1.0 at the equator to about 0.993 at high latitudes. It represents the effect of the flattening of the ellipsoid at the poles, and -- in large-scale maps -- displaces the position of points of high latitude by a visible amount (of the order of a few millimeters) from their position based on a sphere (e = 0). On the other hand, variations in the flattening between one ellipsoid and another (review Section 1.2.1) are not sufficient to have a visible effect.

The most important property of the Mercator projection, from the point of view of navigation, is that a straight line connecting two points on a Mercator chart (Fig. 1.2-39) corresponds to a rhumb line or loxodrome on the ellipsoid.

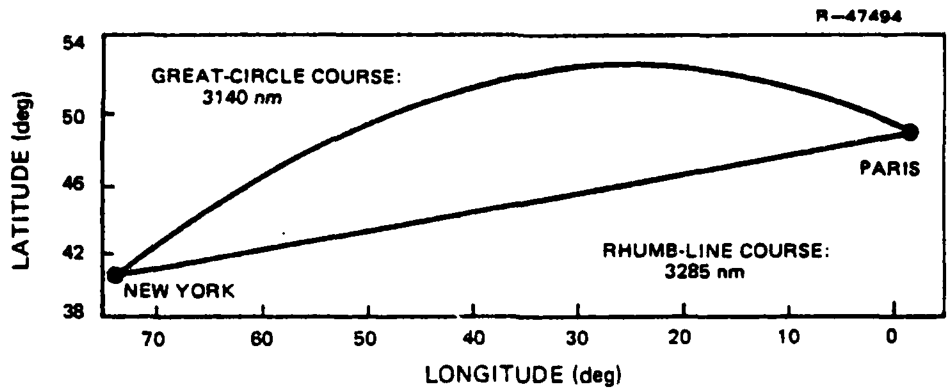


Figure 1.2-39 Rhumb Line on Mercator Projection

This is a curve that crosses all meridians at a constant angle, and is thus the ground track of a vehicle that maintains a constant compass heading. While such a track is, in general, not the shortest path between the two points, it is often preferred because of the convenience of steering a constant heading. It is assumed, of course, that the compass heading referred to is a true heading, as determined by inertial or electronic techniques, or a magnetic heading properly corrected for deviation and variation.

Particularly on a global map, it is quite obvious that the Mercator projection is neither equidistant nor equivalent, since there is a scale expansion with increasing latitude, as shown in Fig. 1.2-40. If the latitude range is limited, though, the distortion is small, and a Mercator projection can be used as a general-purpose map for a region that extends, for example, from the equator to 15 deg North latitude.

Transverse Mercator Projection - Another application of the Mercator principle is based on a projection from the

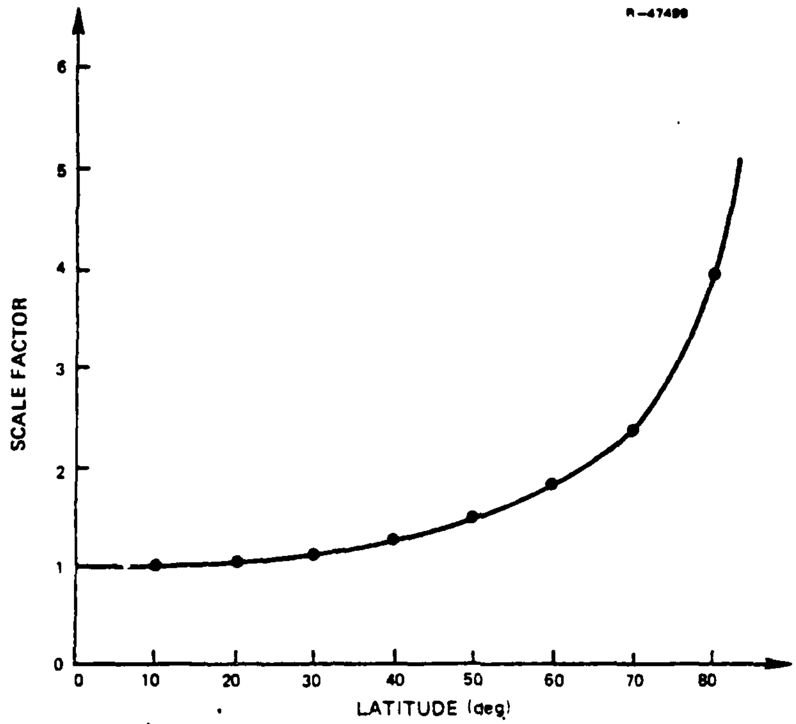
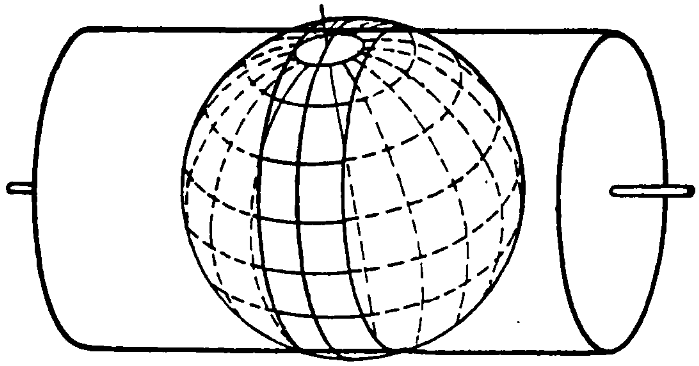


Figure 1.2-40 Scale Expansion in the Mercator Projection

ellipsoid to a horizontal (transverse) cylinder that is tangent along a meridian (Fig. 1.2-41). This projection is useful for maps of regions that are limited in the east-west direction, but of great extent north to south. An example would be a map of South America, or of one of the U.S. states that extends mostly north to south (Maine, Indiana, or Nevada, for example). The equations describing the transverse Mercator projection are quite complicated, but if the earth is approximated by a sphere, then the mapping simplifies to the following form:

$$Y = \frac{1}{2} k \ln \frac{1+\cos\phi \cos\lambda}{1-\cos\phi \cos\lambda} \qquad (1.2\text{-}19)$$

$$X = k \arctan(-\cot\phi \sin\lambda) \qquad (1.2\text{-}20)$$



Note: For the Universal Transverse Mercator (UTM), the radius of the cylinder is slightly less than that of the globe.

Figure 1.2-41 The Transverse Mercator Projection

where

X and Y are the rectangular coordinates of the map

ϕ is the geodetic latitude

λ is the geodetic longitude measured from the central meridian (meridian of tangency)

k is a scale constant

The graticule of the transverse Mercator projection does not consist of straight lines (although on maps of reasonably small regions the curvature of the latitude and longitude lines is not very great), nor does the map have the rhumb line property of the ordinary Mercator projection.

Universal Transverse Mercator - A slight modification of the transverse Mercator projection is of great importance in world mapping, particularly for military applications. The Universal Transverse Mercator (UTM) projection is based on a transverse cylinder of radius slightly less than that of the earth, so that the cylinder cuts the sphere (secant) rather than being tangent to it. The world is divided into 60 zones, each covering 6 deg of longitude. By international agreement, the zones are numbered consecutively toward the east, with Zone 1 lying between 180 deg W and 174 deg W. In each zone, the secant cylinder is symmetric with respect to a central meridian, and the map extends 3° on either side. The latitude coverage is from 80 deg S to 84 deg N.

The fact that the transverse cylinder is a secant, rather than tangent, surface results in a more even spread of scale distortions across the entire zone. The central meridian has a scale factor of 0.9996, rather than 1.0 (see Fig. 1.2-40), and the scale distortion is limited to one part in 2500 throughout the zone.

Another feature of the UTM system is a plane rectangular metric grid superimposed on each zone, with the central meridian given the value 500,000 meters (to avoid negative values). For the same reason, the equator is labelled zero for the northern hemisphere, but is offset by 10,000,000 meters for the southern hemisphere. Grid coordinates (in meters) are often used in place of latitude and longitude.

Lambert Conformal Conic Projection - The Lambert Conformal Conic Projection is the most widely used of the maps based on projection onto a cone. The cone is oriented to be a secant to the ellipsoid, cutting the earth's surface along two parallels of latitude, called the standard parallels. Map

scale is <u>exact</u> along the standard parallels, <u>small</u> between them, and <u>large</u> beyond them (Fig. 1.2-42). For the United States, the standard parallels are generally taken to be 33 deg N and 45 deg N. The meridians on a Lambert projection are straight lines, converging toward a point beyond the limits of the map; the parallels of latitude are concentric circles.

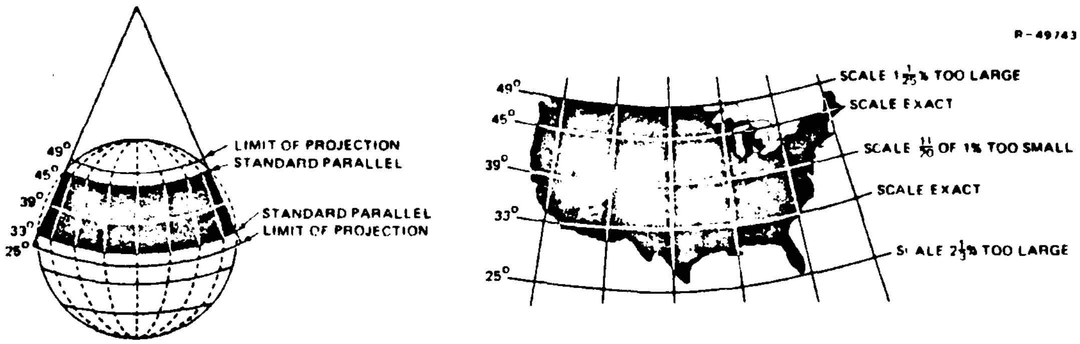


Figure 1.2-42 The Lambert Conformal Conic Projection

The Lambert conformal conic projection is frequently used for state and country maps, aeronautical charts, and radio navigation. Even over an area as large as the United States, the Lambert projection is <u>nearly</u> equidistant and equivalent.

<u>Polar Stereographic Projection</u> - The polar stereographic projection is an example of a map based on projection onto a plane (Fig. 1.2-43). Its graticule consists of meridians (straight lines) radiating outward from the pole, with the parallels of latitude appearing as concentric circles. The projection is used for navigation in the polar regions (within about 20 deg of the pole) as well as for general representation of those regions, since all forms of distortion are small as long as the map does not extend too far from the pole. For

R—49744

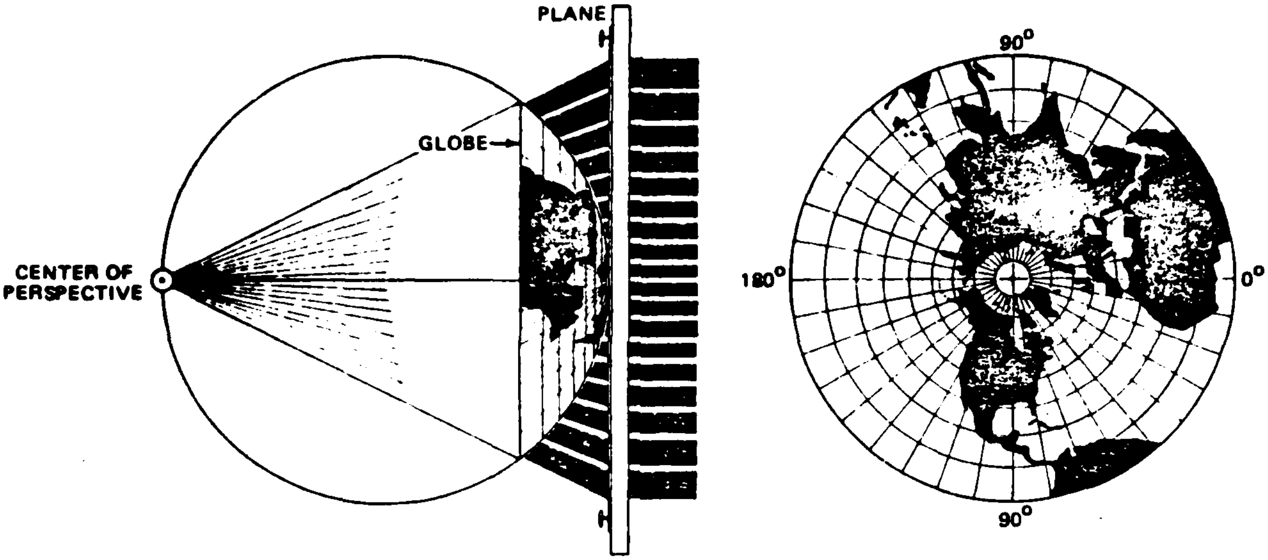


Figure 1.2-43 Polar Stereographic Projection

simplicity, equations for the polar stereographic projection are given only for the spherical case:

$$\rho = 2K \tan (45° - \frac{\phi}{2}) \quad (1.2\text{-}21)$$

$$\theta = \lambda \quad (1.2\text{-}22)$$

$$Y = \rho \cos \theta \quad (1.2\text{-}23)$$

$$X = \rho \sin \theta \quad (1.2\text{-}24)$$

where

ρ, θ are polar coordinates on the map

X, Y are rectangular coordinates on the map

ϕ is geodetic latitude

λ is geodetic longitude

K is a scale factor relating distance on the map to distance on the earth's surface

Equidistant Projections - Equidistant projections are used in situations where accurate representation of distances* (rather than shapes, angles, or areas) is the major criterion. An example of this category is the azimuthal equidistant projection, often used for determining bearing and distance from a specific fixed point, the origin, to other points on the earth's surface. Two applications are global navigation and radio propagation studies.

The azimuthal equidistant projection is not a projection in the geometric sense, but is related to a perspective projection called the gnomonic projection (Fig. 1.2-44a), by an adjustment of the distance scale to achieve the equidistant property (Fig. 1.2-44b).

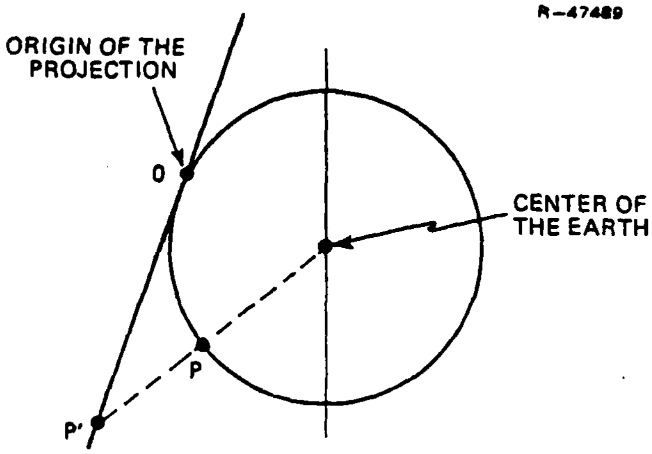
Equivalent Projections - The variety of projections devised to satisfy the criterion of equivalence (equal area) is very great, and many of these -- on a global scale -- obtain equivalence at the expense of considerable distortion in shape and orientation, as shown by the two examples in Fig. 1.2-45. The simplest of the equivalent projections is the Lambert equivalent cylindrical projection, defined (for the spherical case) by the equations

$$Y = K \sin \phi \qquad (1.2\text{-}25)$$

$$X = K \lambda \qquad (1.2\text{-}26)$$

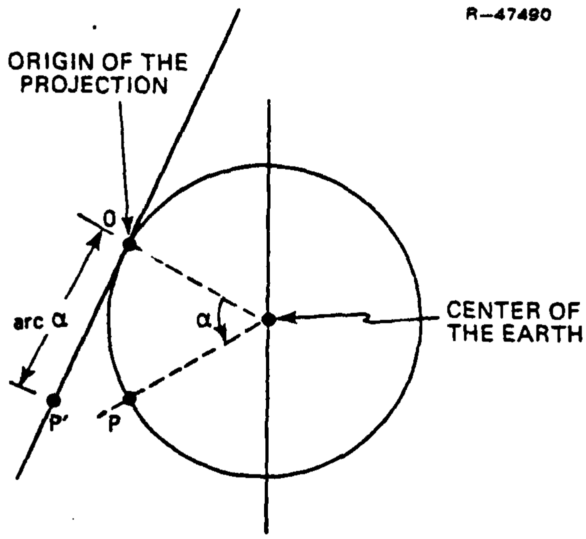
This, like the Mercator projection, is derived from perspective projection onto the cylinder by an alteration of the scale in the north-south direction. For the Mercator projection, the

*Such projections do not reproduce all distances correctly, but only those from one specific point (or, in some cases, from each of two points).



Point P on the globe is projected onto P' on the plane

(a) GNOMONIC PROJECTION

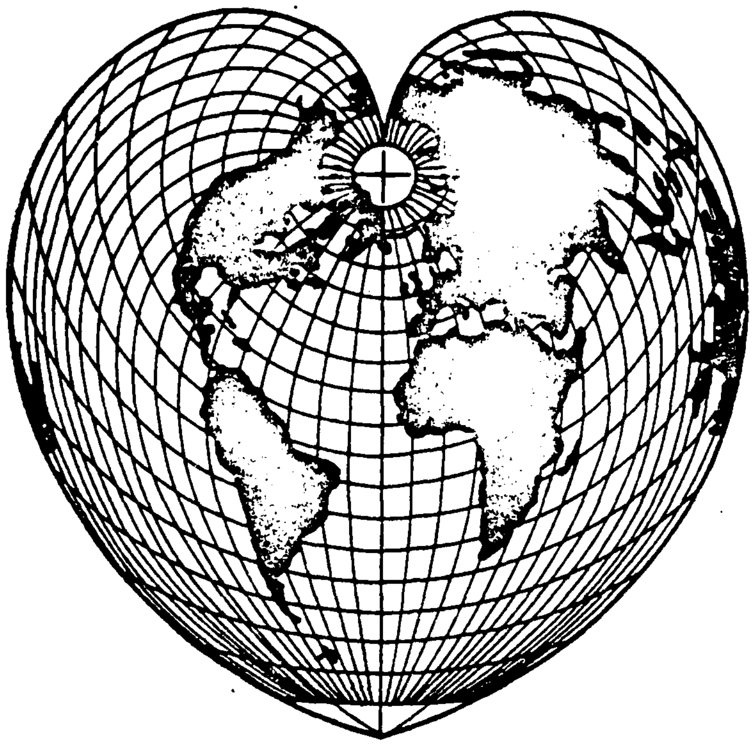


Point P on the globe is mapped into P' on the plane, where the distance OP' on the map is proportional to arc α, the actual distance between O and P on the earth.

(b) AZIMUTHAL EQUIDISTANT PROJECTION

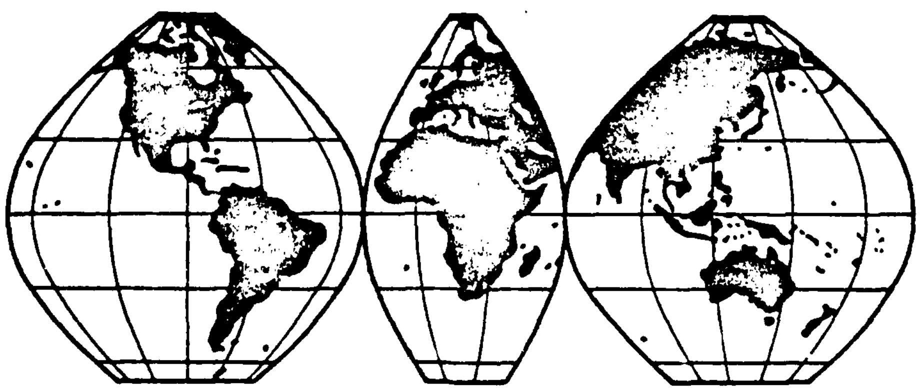
Figure 1.2-44 Gnomonic and Azimuthal Equidistant Projections

R—49745



a) WERNER EQUAL-AREA PROJECTION

R—49746



b) INTERRUPTED SINUSOIDAL EQUAL-AREA PROJECTION

Figure 1.2-45 Examples of Equivalent Projections

scale is <u>expanded</u> at high latitudes; since the process of projection onto the cylinder extends the east-west scale (because the meridians appear to remain at a constant distance apart, rather than converging), the north-south expansion maintains <u>shapes</u> (although <u>areas</u> are distorted). To maintain correct representation of <u>areas</u>, the Lambert equivalent cylindrical projection <u>shrinks</u> the north-south scale in proportion to the expansion of the east-west scale (Fig. 1.2-46). In the vicinity of the equator, this projection is very nearly conformal and equidistant, but distortion is very great at high latitudes.

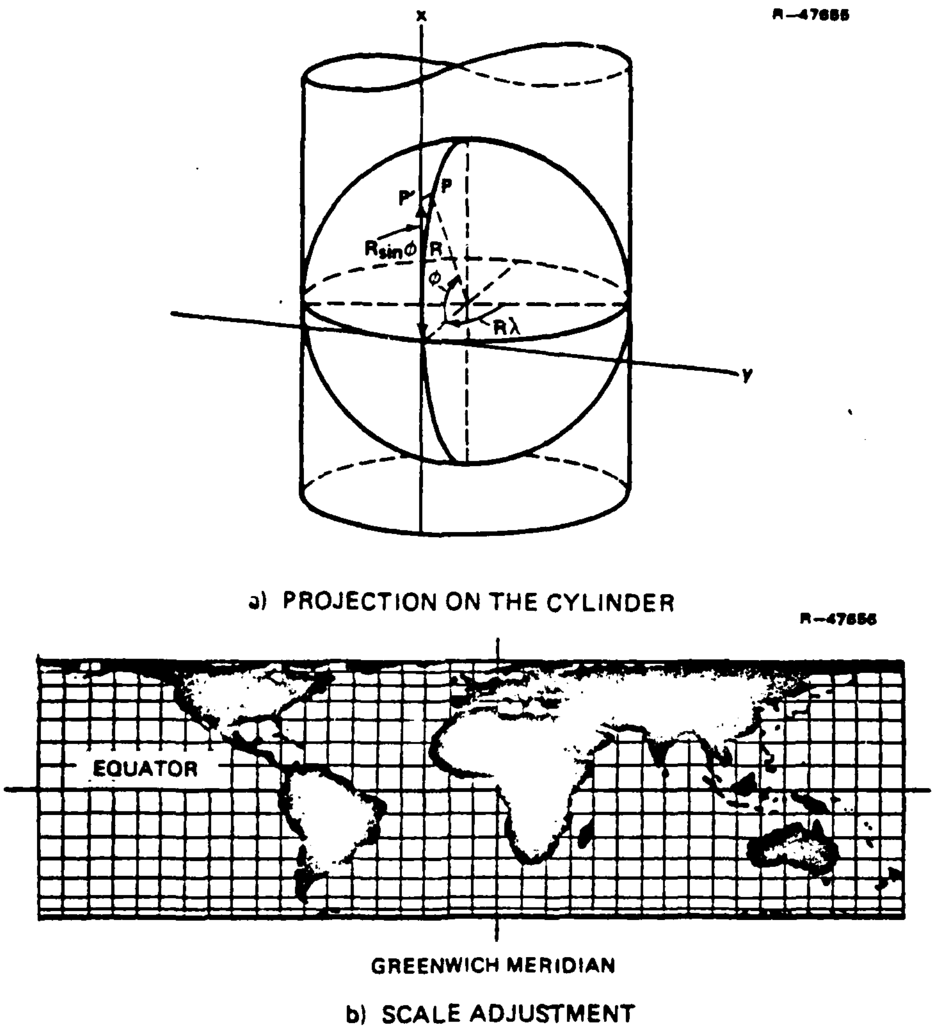


Figure 1.2-46 Lambert Equivalent Cylindrical Projection

1.2.7 Techniques

Two particular techniques of importance in geometric geodesy are presented in this section. These are:

- Inertial Survey Systems (Section 1.2.7.1)
- Radio Interferometry (Section 1.2.7.2).

1.2.7.1 Inertial Survey Systems

Inertial Navigation Systems (INS), to be discussed in more detail in Unit Two, determine a vehicle's velocity and position by integrating measurements of the accelerations experienced by the vehicle.

Inertial systems enjoy widespread use in many weapon systems where high accuracy autonomous navigation or guidance is required. Because of their precision and accuracy, inertial systems are also used as survey tools. Typical inertial surveys involve position determination over a line which can extend fifty kilometers or more. Each such survey line or traverse* is usually completed in several hours -- before instrument errors in the inertial system build up and require the system to be recalibrated. Constant-heading traverses are usually preferred because inertial system calibration factors often deteriorate with azimuth changes. A typical traverse is illustrated in Fig. 1.2-47.

Note that in Fig. 1.2-47, frequent stops of the survey vehicle are indicated. The purpose of these stops is to compare

*The reader may wish to review the concept of traverse in Section 1.2.2.

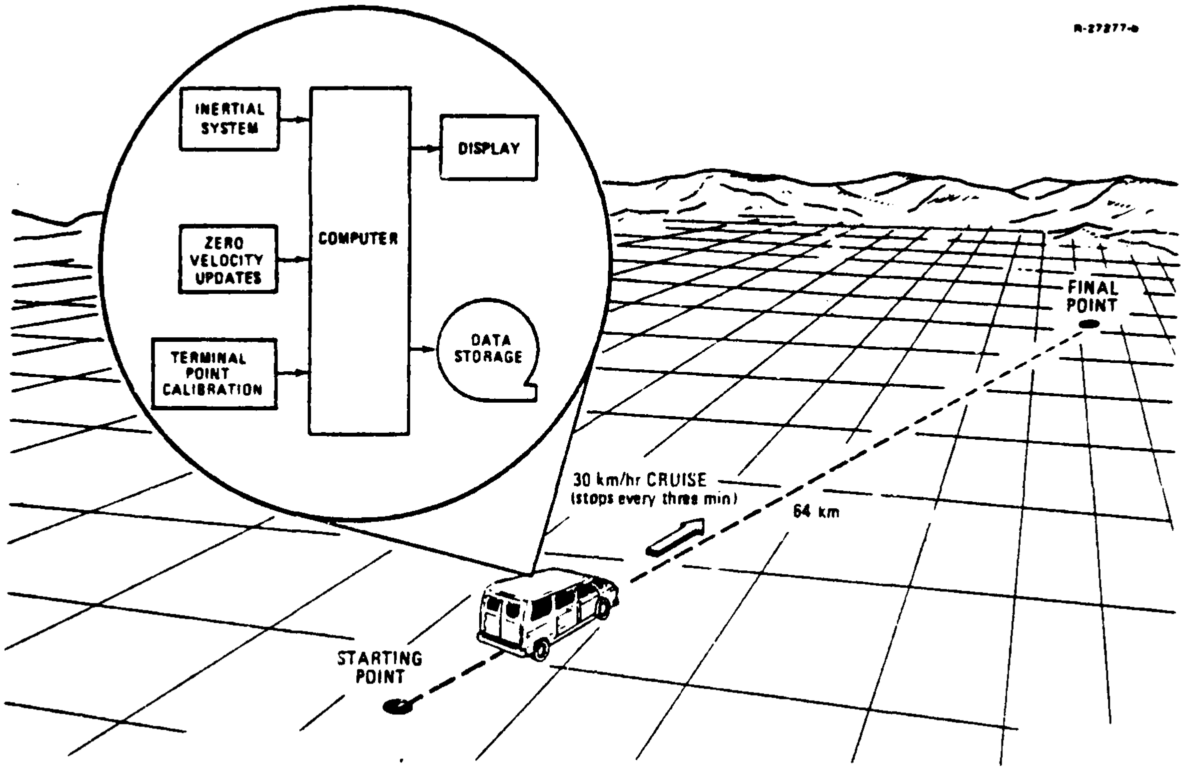


Figure 1.2-47 Inertial Survey Traverse

the inertial system's velocity output to the known (zero) velocity of the vehicle. Any difference is used to trim the inertial system calibration, thus holding the effects of inertial system errors to a very low level. The process of adjusting the inertial system to indicate zero velocity during a vehicle stop is termed zero velocity update or ZUPT.

In addition to automotive vehicles, inertial surveys are often conducted from helicopters, particularly in mountainous or remote areas such as the Arctic. Advantages of helicopter surveys include greater distances covered in the time between ZUPTs and a capability to cover impassable terrain. Disadvantages include the problem of finding landing sites every few minutes, the difficulty of maintaining con-

stant heading, and aircraft upkeep expense. Nonetheless, helicopter-borne inertial surveys provide sufficient accuracy and productivity that several private companies use them in the course of their Alaskan survey operations.

In addition to providing geodetic position, inertial survey systems can also measure the gravity disturbance vector.* The gravity disturbance field is ordinarily an inertial system error source. Unmodeled gravity accelerations are integrated in the same way as vehicle accelerations. As a result, position and velocity errors occur because the gravity accelerations are independent of the vehicle's motion. For survey operations the ZUPT measurements can be used to estimate the gravity disturbance field as well as reduce the effective level of the inertial system instrument errors. These calculations are usually performed after a traverse and typically involve the use of sophisticated mathematical models for the error propagation dynamics of the inertial system.** Post-mission data processing is usually accomplished with algorithms that take advantage of the different behavior of gravity and the inertial system errors over the length of the traverse.

Because inertial system errors grow with time, the accuracy of an inertial survey deteriorates from the time of initial point calibration to the end of the traverse. If a recalibration at the final point of the traverse is used in combination with post-mission processing, the largest errors occur near the survey midpoint. Such a case is illustrated in Fig. 1.2-48 for a simulated survey 64 km in length with ZUPTs

---

*The gravity disturbance (discussed in Chapter Three) is the difference between the true (measured) value of gravity at a point, and the (approximate) value calculated from a gravity model.

**Inertial system error dynamics is discussed in Unit Two.

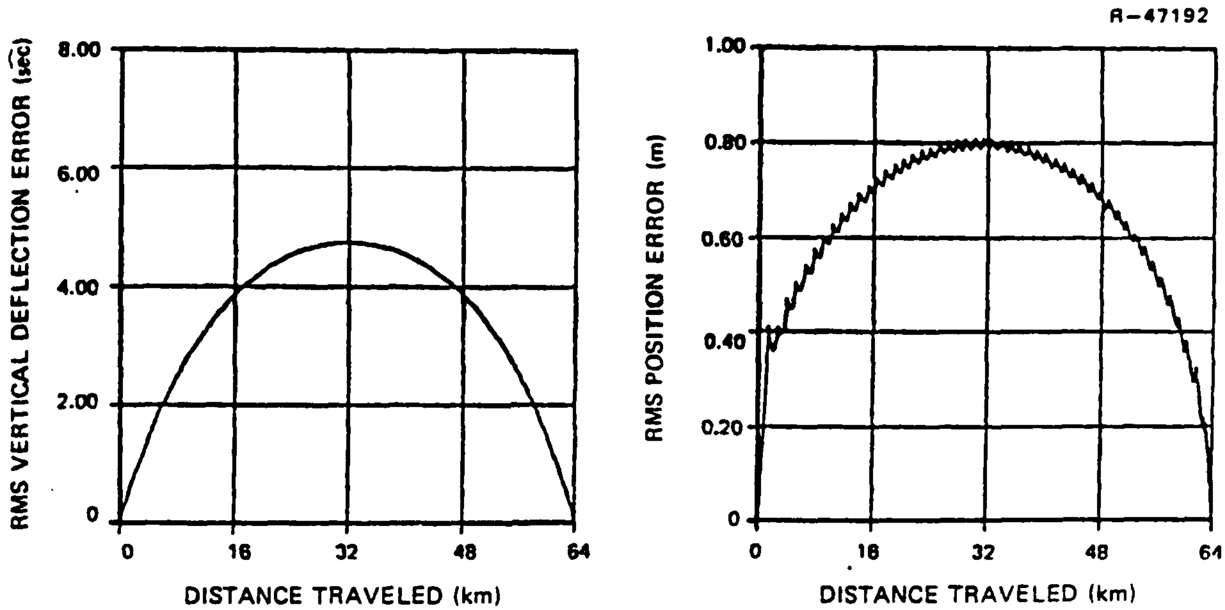


Figure 1.2-48 Survey Error vs Distance Along Traverse

every three min. Precise position and vertical deflection calibration of the inertial system is assumed to be performed at the terminal points of the traverse. The effect of the ZUPT measurements is readily apparent in the position error profile. Note the steep rise in position error from the start of the survey until the first ZUPT. If it were not for the ZUPT measurements, survey error would continue to grow rapidly until, within a few kilometers of the starting point, the system output became unusable.

Current inertial survey system hardware and software can achieve accuracies of about 0.5 m for lateral position and 30 - 40 cm in height over traverse distances of about 75 km. Gravity and deflection of the vertical determinations can be accurate to 2 mgal and 1.5 $\widehat{sec}$, respectively. Users of helicopter-based systems, like the Geodetic Survey of Canada, are achieving comparable horizontal position accuracies and slight-

ly less accuracy vertically (because of rapid vertical excursions of the helicopter). The mission distances by helicopter are appreciably greater. Typically, they range from 150 to 200 km.

Inertial survey systems in wide use today are variants of the Litton LN-15 aircraft inertial navigation system. Other inertial survey systems are marketed by Ferranti (Scotland) and Honeywell. More modern inertial components, particularly improved gyros, could improve performance appreciably, particularly by permitting the time interval between ZUPTs to be increased. The Honeywell system may represent a step in this direction insofar as it incorporates a more advanced gyro and inertial platform technology (electrostatically suspended gyros). It is clear that the future trend in inertial surveying is toward the incorporation of high quality inertial components and software into systems capable of higher accuracy, increased ZUPT intervals, and longer traverse times.

1.2.7.2 Radio Interferometry

Introduction - Interferometry is a distance measurement technique which takes advantage of the wave-like nature of radiation to achieve very high accuracies. In particular, signals which propagate over different, but similar, paths can be made to nearly cancel (or interfere with) each other. Interferometric measurement systems are designed so that the residual signal, after cancellation, provides information about the propagation-path difference. Because very small propagation distance differences can be measured accurately, interferometry is of particular interest to the geodesist as a precision distance measuring tool.

Interferometric methods were first applied to visible light and used to determine differences in optical pathlengths of the order of hundreds of Ångstroms.* In early optical systems the measurement was a series of light and dark areas on a screen (<u>fringes</u>). The fringes are projections of the interference pattern which results when two light beams originating from the same source, but traversing slightly different optical paths, are superposed. The distance between fringes is proportional to the wavelength of the light. Highly accurate relative velocity measurements can be made by counting fringe rates (i.e., number of interference maxima or minima passing an observation point in unit time). This technique allowed the velocity of light to be measured with sufficient precision in the classic Michaelson-Morley experiment to establish that the speed of light is constant, regardless of the speed of the source or observer.

Radio interferometry, like its optical counterpart, measures the propagation pathlength differences of two radio signals from the same source, as illustrated schematically in Fig. 1.2-49.

The extra time, Δt, that the signal takes to reach receiver No. 1 defines the additional path length $c\Delta t$, where c is the speed of light. In terms of the baseline distance between the two receivers, b, and the look-angles β_1 and β_2, the time delay is given by

$$\Delta t = \frac{b}{c} \frac{\cos \frac{\beta_1+\beta_2}{2}}{\cos \frac{\beta_1-\beta_2}{2}} \tag{1.2-27}$$

*Ångstrom Unit = 10^{-10} m.

Comparison of the two received signals allows the delay time, Δt, to be ascertained. This is done by time shifting one signal with respect to the other and observing when the two signals most closely match or correlate. The delay time information can be used to determine the length of b when the antenna pointing angles, β_1 and β_2, are known. Note that Δt is a differential measurement; neither the distance from the receiver to the radio source nor the propagation characteristics over the equal portions of the paths need to be known. If the source is far away compared to the baseline length, β_1 is essentially equal to β_2 and Eq. (1.2-27) simplifies to

$$\Delta t = \frac{b}{c} \cos \beta_1 \qquad (1.2\text{-}28)$$

This approximation is particularly appropriate when stellar objects are used as radio sources. In such cases, when the value of b is known, Eq. 1.2-28 can be used to determine the celestial position of a radio source to much higher accuracy than antenna resolvers would allow. Artificial satellites can also be used as radio sources, and several schemes have been proposed using signals from satellites to determine geodetic distances among receiving stations in a ground network. Accuracy is expected to be of the order of a few cm.

Practical Considerations[†] - Of course Fig. 1.2-49 and Eq. 1.2-27 are quite simplified descriptions. Operational radio interferometry is conducted with advanced radio telescope equipment and typically involves the support of large university communities. Since the baseline is located on the earth, it continually changes position with respect to an extraterrestrial radio source. In addition to accounting for the radio

(†)This section contains material at a more advanced level than the rest of the text.

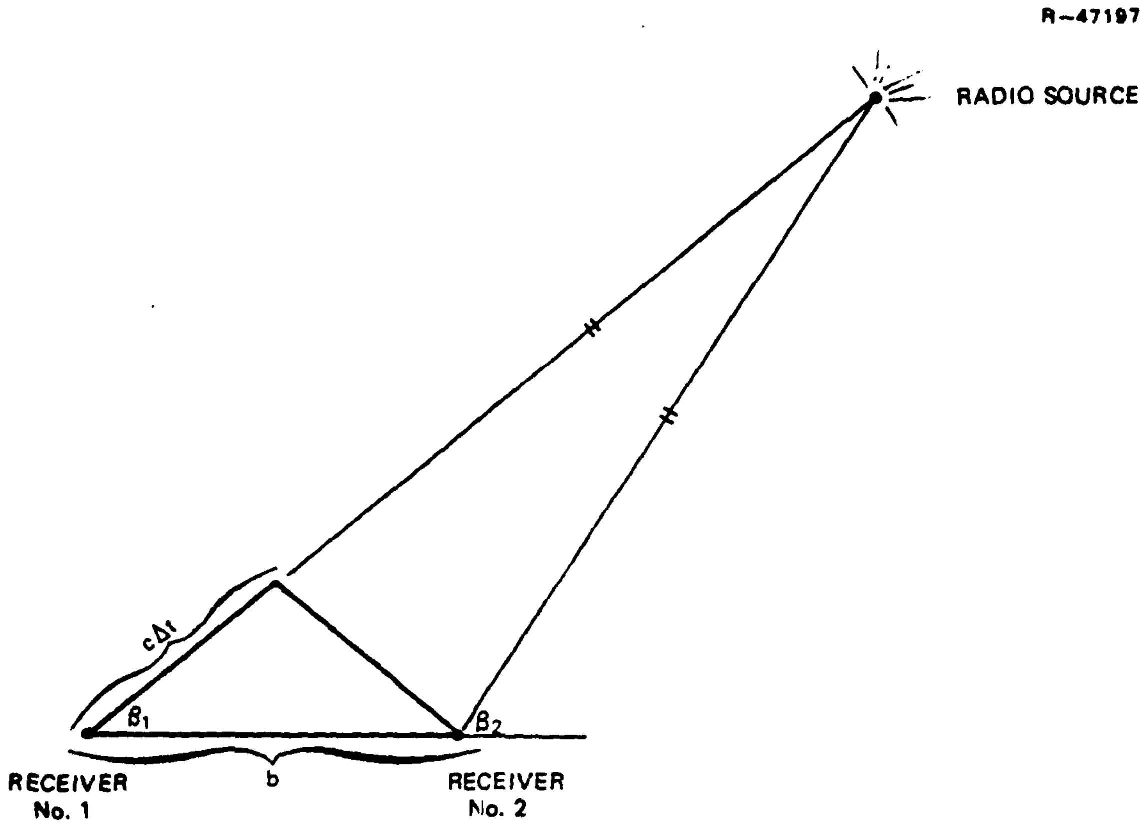


Figure 1.2-49 Simplified Radio Interferometer Measurement Geometry

source's position in the sky (both right ascension and declination, rather than a single look angle such as β_1), corrections must be made for anomalous time delays caused by atmospheric differences along the two propagation paths. Additional corrections for instrumentation-caused propagation delay differences are required as well. Finally, there is an ambiguity in the correct amount of phase delay which brings the output signals from the two receivers into coincidence. The ambiguity is readily apparent when a single frequency transmission is considered. If the output of the first receiver is

$$S_1(t) = \cos 2\pi ft \qquad (1.2\text{-}29)$$

then the output of receiver two will be

$$S_2(t) = \cos 2\pi f(t+\Delta t) \qquad (1.2\text{-}30)$$

where f is the frequency and Δt is the propagation delay illustrated in Fig. 1.2-49.

If the received signal S_1 is shifted by an amount τ until it matches or correlates with S_2, i.e.

$$S_2 = S_1(t+\tau) = \cos 2\pi f(t+\tau) \qquad (1.2\text{-}31)$$

an infinite number of solutions for τ are possible, namely

$$\Delta t = \tau \pm \frac{n}{f} \qquad n = 0,1,2,... \qquad (1.2\text{-}32)$$

Often, when the baseline is small, the phase ambiguity number, n, can be determined by allowable limits on the range of Δt.

For example, suppose it is known that the differential propagation length, $c\Delta t$, is between 20 and 30 cm, corresponding to a propagation delay time difference* in the range $0.67 < \Delta t < 1.0$ nsec. Let the frequency be 5 GHz, (5×10^9 Hz) and suppose data analysis shows the smallest value of τ which can correlate the two receiver outputs to be 0.23 nsec. Using Eq. 1.2-32,

$$\Delta t = 0.23 + 0.2n \text{ nsec} \qquad (1.2\text{-}33)$$

Since only one value of n leads to a result that falls within the allowable range of Δt, n is three and the actual differential propagation length is 24.9 cm.

---

*c = 30 cm/nsec (1 nsec = 10^{-9} sec).

However, many radio interferometer observations do not permit such simple resolution of the phase ambiguity. Often, too, the change in n over the time interval of the observations is not easy to ascertain. The value of n may be different for observations of different sources or even for observations of an individual source made at different times (because of variations in baseline orientation, the atmosphere, or instrumentation delay). A variety of techniques are used to surmount this difficulty, one being to make observations at several frequencies. Advanced signal processing techniques are used to take advantage of the fact that each measurement spans a finite rather than infinitesimal frequency bandwidth. This approach is referred to as group delay measurement.

Systems - Two types of interferometer systems are of particular interest for application to geodetic distance determination. One, often referred to as conventional, involves an electrical connection between the separate antennas. The interferometric phase delay is measured in real time. In the other type, the so-called very-long-baseline interferometer, (VLBI), there is no real-time connection between the antennas. The received signals are tape-recorded simultaneously, but independently, at the two sites and the recordings are later cross-correlated to determine the interferometric observables. The most important features of both types of interferometer systems are described below.

Figure 1.2-50 shows, in simplified form, a typical conventional interferometer with two antenna-receiver systems. At each antenna the radio-frequency (RF) signal received from the source being observed is converted to a lower intermediate frequency (IF) by mixing with a local-oscillator (LO) signal. The LO signals are supplied to the mixers at both antennas via transmission lines from the centrally located oscillator. The

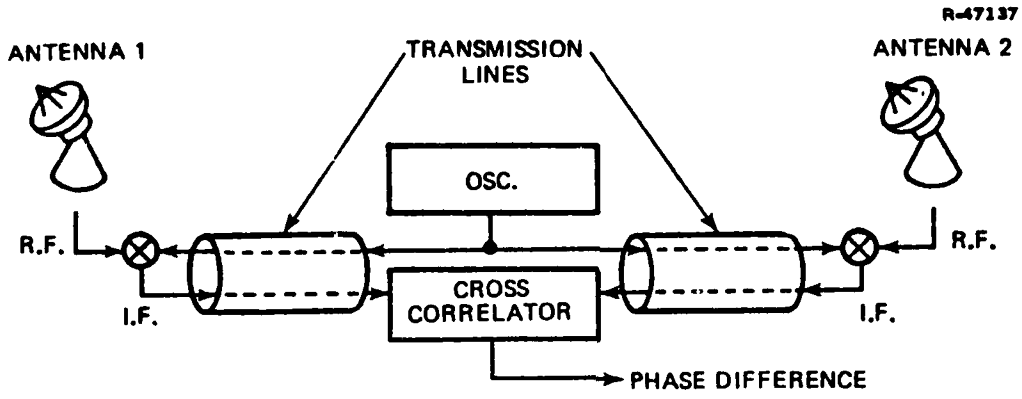


Figure 1.2-50 Conventional, Connected-Element Radio Interferometer

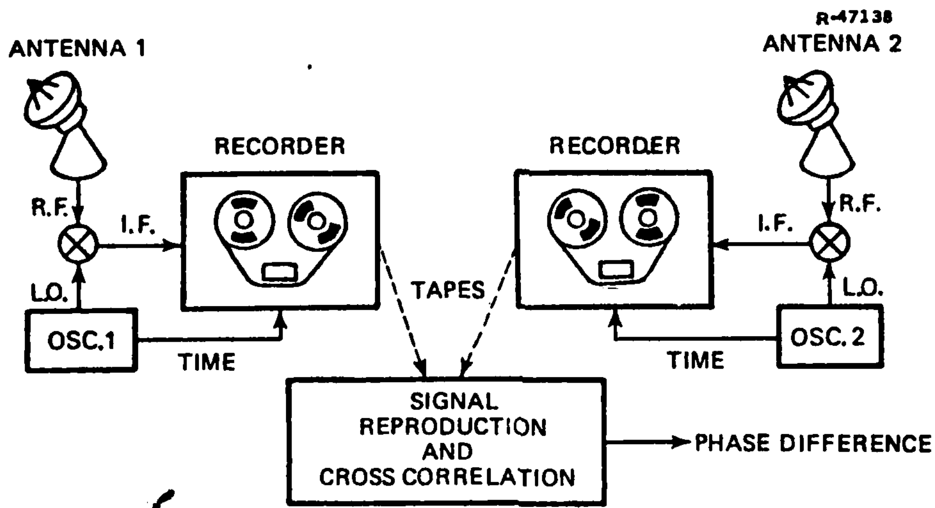


Figure 1.2-51 Very-Long-Baseline Interferometer

IF signals are carried by similar lines back to the central station where the interferometric phase, equal to the difference between the RF signal phases, is determined by cross-correlation of the two IF signals. Ideally, the electrical

path lengths of the transmission lines from the central oscillator to the two antennas' mixers, and the phase delays of the two sets of receiving electronics, are equal; any differences not accounted for will introduce errors in the observed interferometric phase. In current practice, these phase errors can be reduced to the level of one deg at a frequency of 5 GHz, or the equivalent of 0.2 mm of path-length error for a 5-km baseline interferometer.

In the VLBI, the LO signal used for the RF to IF conversion at each antenna is derived from an independent frequency standard (see Fig. 1.2-51). At each site the IF signal is tape recorded with a reference time base derived from the same standard. Tapes recorded simultaneously at the two antenna sites are later replayed at a processing station, where the reproduced signals are cross-correlated to determine the interferometric phase and related observables.

The advantage of substituting independent frequency standards and tape recorders for real-time signal transmission links is an economic one: once the need for a real-time connection between the ends of the baseline is eliminated, baseline lengths of thousands of kilometers become practical. At present, the main disadvantages of VLBI are that: (a) the IF bandwidth limitation set by the tape recorders may be more stringent than the corresponding limitation of a real-time transmission medium, and (b) very high stabilities are demanded of the frequency standards.

Accuracy - Currently operational radio interferometric systems can measure the angular position of stellar radio sources to an accuracy of 0.05 $\widehat{\text{sec}}$. Relative positions of sources with small angular spacings can be determined to 10^{-3} $\widehat{\text{sec}}$. Distance measurement accuracy for short baselines (sev-

eral kilometers) has been demonstrated to be better than one centimeter (by comparison with other measurement techniques). Similar comparisons for very long baselines (>5000 km) show current state-of-the-art accuracies of the order of one meter or better. The internal self-consistancy and repeatability of the VLBI measurements is about 10 cm.

Applications - In addition to obvious uses of interferometry for the determination of geodetic distances, other possible applications include measurement of the earth's crustal motion and monitoring of earthquake fault displacements. Radio interferometry has also been used to determine changes in universal time (UT0), measure polar motion and precession, and detect solid earth tides. Scientific applications range from tests of general relativity to the calibration of other precise position measuring systems such as those using radio navigation satellites.

CHAPTER THREE
PHYSICAL GEODESY

Physical geodesy is concerned with the gravity field of the earth. Certain basic concepts -- vector and scalar fields, potential, gravity, and gravitation -- are fundamental to the material covered in this chapter. While familiarity with these concepts is assumed (from previous course work in physics and mechanics), a brief review is provided for the reader's convenience in Appendix A.

1.3.1 The Geoid

If the earth were entirely fluid, its physical surface would be a surface of constant gravity potential. The water-covered four-fifths of the planet should, therefore, obey this rule, and the surface agreeing with mean sea level (assuming the effects of tides, wind, currents, etc. to be averaged out) is a surface of constant potential called the geoid. For land areas, the geoid is defined in a less direct manner, by imagining that the sea level surface is extended beneath the land, in imaginary canals or conduits, with the water permitted to rise to conform to a surface of constant potential (Fig. 1.3-1). This basic definition implies that some of the mass of the earth, in continental areas, lies above the geoid. For theoretical work in geodesy, this is sometimes undesirable. Thus, other surfaces, related to the geoid, are defined in such a way that the surface encloses the entire mass of the earth.

The geoid is an irregular surface for which no convenient mathematical expression exists, although it may be approximated in various ways. It is defined by its relation to a

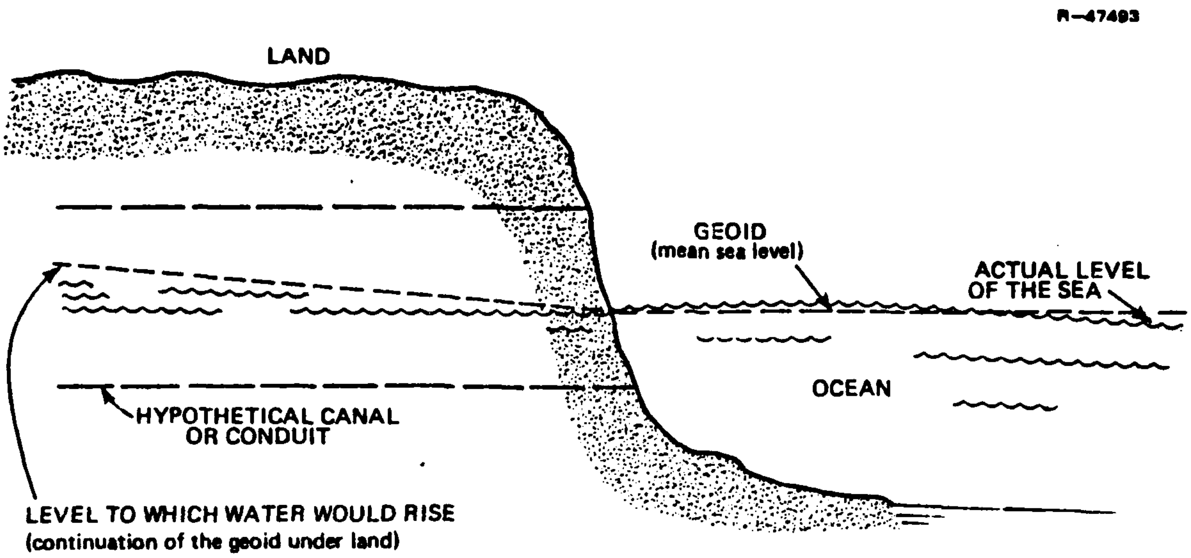


Figure 1.3-1 Concept of the Geoid

particular ellipsoid, as shown in Fig. 1.3-2, in terms of the geoid height, or undulation, at every point. One way of describing the geoid is by the use of a map, like Fig. 1.3-3, showing contours of equal geoid height. Another important relation between the geoid and the ellipsoid, also shown in Fig. 1.3-2, is the deflection of the vertical. This is the angle between the normal to the geoid and the normal to the ellipsoid.

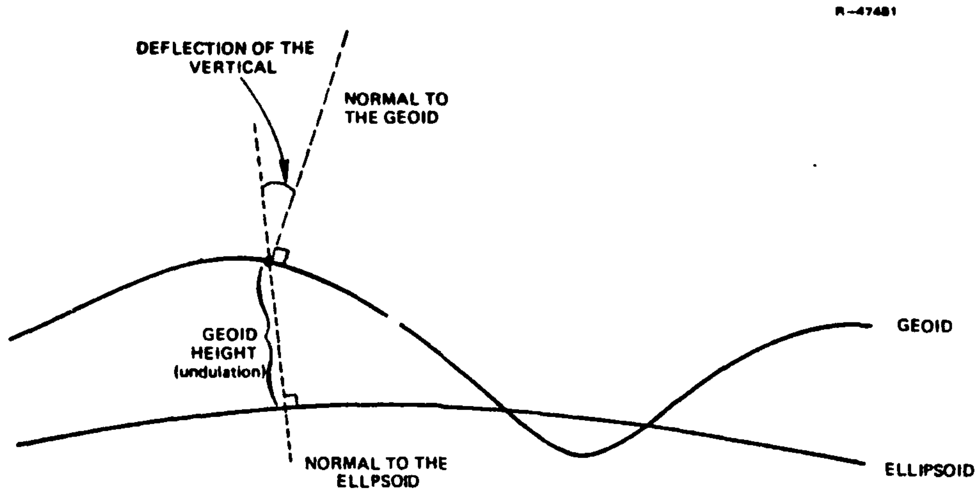


Figure 1.3-2 Relation of Geoid and Ellipsoid

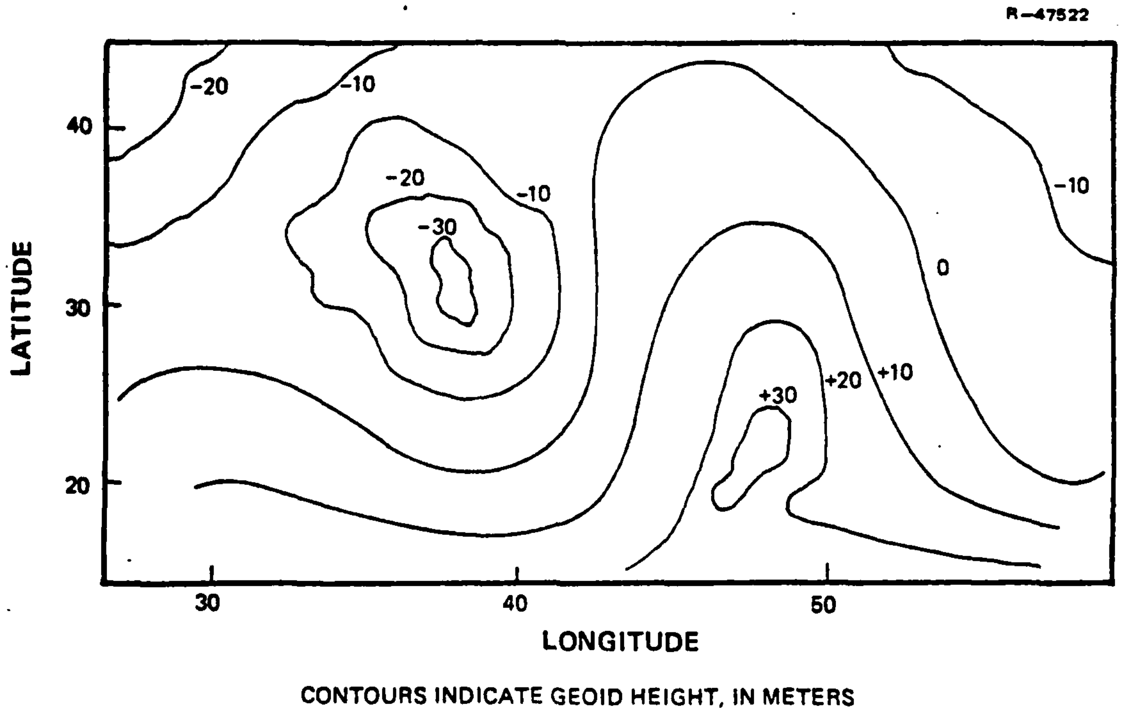


Figure 1.3-3 Example of Geoid Contour Map

1.3.2 Gravity

Gravity, at any point on the earth's surface, is the resultant of two accelerations acting in opposite directions:

- Gravitation, the integrated effect of the attraction of all of the mass making up the earth
- Centrifugal, the outward-tending acceleration caused by rotation of the earth.

As an acceleration, gravity is measured in the basic units of meters per second per second (m/sec^2) and has the approximate surface value of 9.8 m/sec^2. Other units frequently used in gravity studies are summarized in Table 1.3-1.

TABLE 1.3-1

GRAVITY UNITS

UNIT	ABBREVIATION	CONVERSION
gal	gal	1 gal = 10^{-2} m/sec^2
milligal	mgal	1 mgal = 10^{-3} gal = 10^{-5} m/sec^2
microgal	μgal	1 μgal = 10^{-6} gal = 10^{-8} m/sec^2
gravity unit	gu	1 gu = 10^{-6} m/sec^2 = 0.1 mgal

Since the value of gravity varies considerably from point to point on the earth's surface, it is of vital importance in many military applications to be able to predict, or model, the value of gravity to be expected at any given location. As a first approximation, a formula may be used to predict sea-level gravity which is based on the assumption that the ellipsoid (or other spheroid) approximating the earth is,

itself, a surface of constant potential. Such a formula, called a normal gravity formula, is associated with a particular ellipsoid. Many such formulas have been used at various times.

For example, when the Helmert Ellipsoid (a = 6378200 m; f = 1/298.2) was widely used in the early part of this century, the corresponding normal gravity formula was

$$\gamma = 978030(1+0.005302 \sin^2\phi - 0.000007 \sin^2 2\phi) \qquad (1.3\text{-}1)$$

where

γ is the normal gravity, in mgal

ϕ is the geodetic latitude

A formula adopted in 1917 by the U.S. Coast and Geodetic Survey was based on modified values for the semi-major axis and the flattening. It is of the same form, but with slightly altered coefficients:

$$\gamma = 978039(1+0.005294 \sin^2\phi - 0.0000007 \sin^2 2\phi) \qquad (1.3\text{-}2)$$

In 1930, an ellipsoid and an associated gravity formula were adopted for international use, known as the International Ellipsoid and the International Gravity Formula. The ellipsoid parameters are

$$a = 6378388 \text{ m} \qquad (1.3\text{-}3)$$

$$f = 1/297.0 \qquad (1.3\text{-}4)$$

and the gravity formula

$$\gamma = 978049(1+0.0052884 \sin^2\phi - 0.0000059 \sin^2 2\phi) \qquad (1.3\text{-}5)$$

This was the standard formula until the adoption of the current international standard, the Geodetic Reference System 1967 (GRS 67), based on the ellipsoid parameters

$$a = 6378160 \text{ m} \tag{1.3-6}$$

$$f = 1/298.247 \tag{1.3-7}$$

The gravity formula for GRS 67 is presented in the following form:

$$\gamma = 978031.84558 \, \frac{1 + (0.00193166338321) \sin^2 \phi}{[1-(0.00669460532856) \sin^2 \phi]^{1/2}} \tag{1.3-8}$$

for which the following approximations are often used:

$$\gamma = 978031.85 \, (1 + 0.005278895 \sin^2\phi + 0.000023462 \sin^4\phi) \tag{1.3-9}$$

with a maximum error of 0.004 mgal, and

$$\gamma = 978031.8 \, (1 + 0.0053024 \sin^2\phi - 0.0000059 \sin^2 2\phi) \tag{1.3-10}$$

which is accurate to within about 0.1 mgal.

It is seen from any of these formulas that the value of gravity increases by about 0.5% from the equator to the pole.

In the remainder of this section, three approaches to the more precise modeling of gravity will be considered.

Global Gravity Models - Global models that predict gravity at any point on the surface, as well as gravitation at points in near-earth space, are of paramount importance. Such models express the scalar potential of gravity as a function of geocentric spherical coordinates in an earth-fixed coordinate system. From the potential, the gravity vector is obtained by differentiation (using the gradient operator), while the component of gravity in any particular direction is given by the directional derivative in that direction. Models for the potential of gravity (usually called geopotential) are of the general form

$$W(r,\phi',\lambda) = V(r,\phi',\lambda) + \frac{1}{2}\,\omega^2 r^2 \cos^2\phi' \qquad (1.3\text{-}11)$$

where

W is the potential of gravity

V is the potential of gravitation

r,ϕ',λ are the geocentric spherical coordinates (radius vector, latitude, longitude)

ω is the earth's angular velocity of rotation (about 7.3×10^{-5} rad/sec)

Depending on whether it is gravity or gravitation being modeled, the second term on the right side of Eq. (1.3-11) is included or left out. The geocentric latitude, ϕ', is written with a prime to avoid confusion with geodetic latitude, ϕ (review Section 1.2.1). The function V is written in the form of a spherical harmonic expression, which is an extension of the concept of a Fourier series (harmonic expansion) to the two-dimensional surface of a sphere:

$$V = \frac{GM}{r} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left(\frac{R}{r}\right)^n P_{nm}(\sin\phi')[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda] \qquad (1.3\text{-}12)$$

In this formula, V is the potential of gravitation; r, ϕ', and λ are geocentric spherical coordinates; R is the equatorial radius of the earth, and r is the distance from the earth's center. The physical constants G, M, and R are described in Table 1.3-2. The set of functions $P_{nm}(x)$ are known as associated Legendre functions, the properties of which are reviewed in Appendix B. The set of constants C_{nm} and S_{nm} are the spherical harmonic coefficients (informally referred to as "the Cs and Ss") that specify the particular model. Equation (1.3-12) is thus seen to specify the geopotential as a combination of individual components of the form

$$V = \frac{GM}{r} (V_{nm1} + V_{nm2}) \tag{1.3-13}$$

with

$$V_{nm1}(r,\phi',\lambda) = \left(\frac{R}{r}\right)^n P_{nm}(\sin\phi') \cos m\lambda \tag{1.3-14a}$$

and

$$V_{nm2}(r,\phi',\lambda) = \left(\frac{R}{r}\right)^n P_{nm}(\sin\phi') \sin m\lambda \tag{1.3-14b}$$

where the index n is the degree and the index m is the order. Each component is called a spherical harmonic and is the product of two parts; for example,

$$V_{nm1}(r,\phi',\lambda) = \left(\frac{R}{r}\right)^n Y_{nm1}(\phi',\lambda) \tag{1.3-15}$$

where

$$Y_{nm1}(\phi',\lambda) = P_{nm}(\sin\phi') \cos m\lambda \tag{1.3-16}$$

TABLE 1.3-2
PHYSICAL CONSTANTS IN THE EXPRESSION
FOR THE GEOPOTENTIAL

SYMBOL	NAME	APPROXIMATE VALUE (SI UNITS)
G	Universal constant of gravitation	$6.67\times10^{-11}\ m^3/kg\ sec^2$
M	Mass of the earth	5.97×10^{24} kg
GM	Geocentric gravitational constant	$3.986\times10^{14}\ m^3/sec^2$
R	Earth's equatorial radius	6.38×10^{6} m
ω	Earth's angular rotation rate	7.292×10^{-5} rad/sec

called a <u>surface spherical harmonic</u>, describes the pattern of variation over the surface of the sphere, and the $\left(\frac{R}{r}\right)^n$ term indicates how this particular contribution to the geopotential <u>decreases</u> as r <u>increases</u>. The surface spherical harmonics are of three kinds, as shown in Table 1.3-3 and Fig. 1.3-4.

TABLE 1.3-3
TYPES OF SURFACE SPHERICAL HARMONICS

TYPE	INDEX VALUES	ILLUSTRATION	COMMENTS
Zonal	m=0	Fig. 1.3-4A	No longitude variation; P_{no} is usually written as P_n (Legendre polynomial); n zeroes along a meridian
Sectoral	m=n	Fig. 1.3-4B	No zeroes along a meridian; 2n zeroes around a parallel of latitude
Tesseral	$m\neq0$ $m\neq n$	Fig. 1.3-4C	2m zeroes around a parallel of latitude; n-m zeroes along a meridian

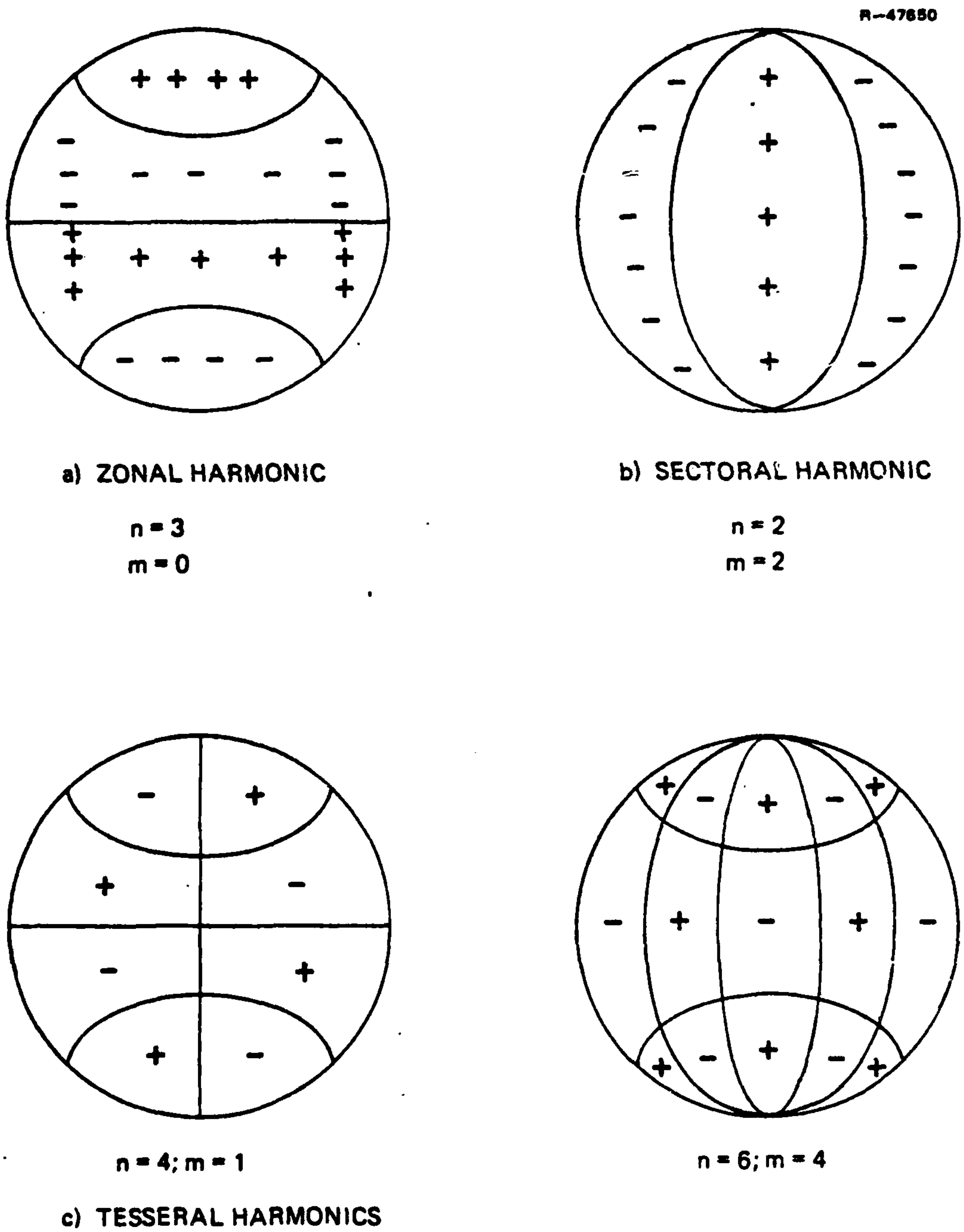


Figure 1.3-4 Examples of Surface Spherical Harmonics

The + signs in Fig. 1.3-4 designate areas of the surface in which the surface spherical harmonic is positive; the - signs, areas where the function is negative. On the lines separating such areas, the harmonic function is equal to zero. For example, in Fig. 1.3-4a, a point moving along the central meridian from north (top) to south (bottom) would traverse a positive region, a zero line, a negative region, a zero line (the equator), a positive region, a zero line, and a negative region. As a second example, a point moving from left to right along the equator of Fig. 1.3-4b successively traverses regions of negative, positive, and then negative values.

In theory, the C_{nm} and S_{nm} coefficients could be determined exactly, by mathematical formulas, if either

- The density were known at every point within the earth
- The gravity were known at every point on the surface of the earth.

In practice, a limited set of coefficients is determined from available surface measurements and/or by analyzing the effects of the earth's gravity field on the orbits of near-earth satellites (which are described at greater length in Chapter Four).

A simpler model of gravity and gravitation ignores the variation with longitude, retaining only terms with m=0 in Eq. (1.3-12). Then the potential becomes

$$V(r,\phi') = \frac{GM}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R}{r}\right)^n P_n(\sin\phi')\right] \qquad (1.3\text{-}17)$$

where

$J_n = -C_{no}$

$P_n = P_{no}$ (the Legendre polynomial of degree n)

Regional Gravity Models - From a review of Fig. 1.3-4, it should be evident that harmonic terms with very high values of n and m are necessary to describe variations in gravity on a scale much smaller than continental. The present state of the art cannot produce spherical harmonic models of high enough order to describe the fine structure of the gravity field at the level required, for example, for a missile launch area. An example of an approach to detailed modeling of gravity within a small region (a few degrees or less) is the point mass model. This model represents gravity as the sum of two parts:

- Gravity as predicted by the normal gravity formula (Eq. 1.3-8) or perhaps by an available spherical harmonic model
- Local effects as modeled by a number of hypothetical point masses of arbitrary mass and location.

The locations and mass values for the point masses are determined by requiring a best fit to gravity actually measured at a large number of control points within the region of interest. Point mass models are discussed at further length in Section 2.3.4 of Unit Two.

Local Gravity Modeling - At any particular point, the actual measured gravity can differ by a significant amount (hundreds of mgals) from values computed by gravity formulas. These differences are referred to as gravity anomalies or gravity disturbances, which may be defined informally as the difference between actual and computed gravity at a point. The magnitude and direction of the gravity vector may deviate significantly (up to a minute of arc, or more) from what is

computed by finding the difference between the gravity vector normal to the geoid (the actual direction of the gravity vector) and the vector normal to the computational surface (the ellipsoid). This difference in direction is called the deflection of the vertical (Fig. 1.3-5). It is identified by its meridian, (north-south) component (ξ), and its prime vertical (east-west) component (η), measured in seconds of arc.

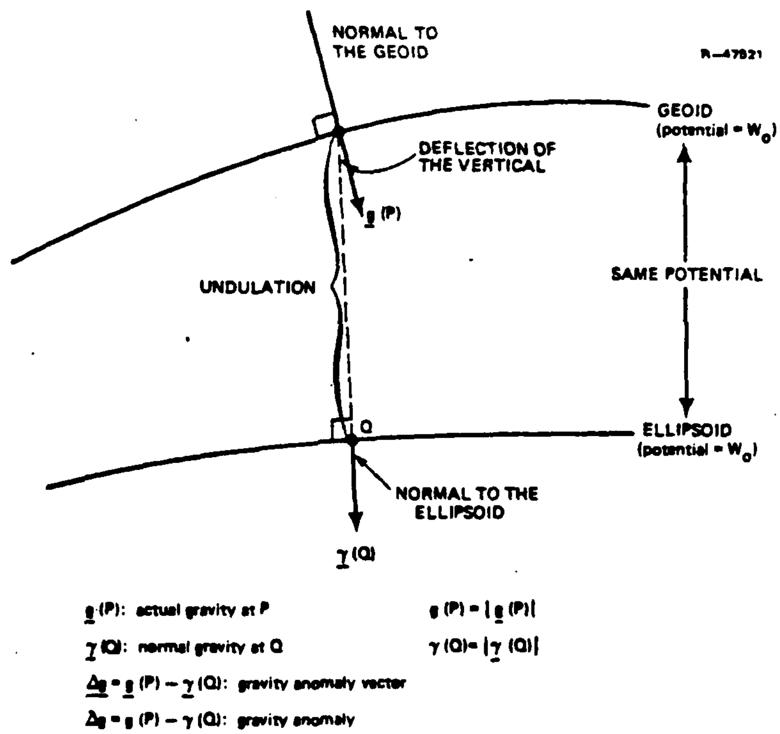


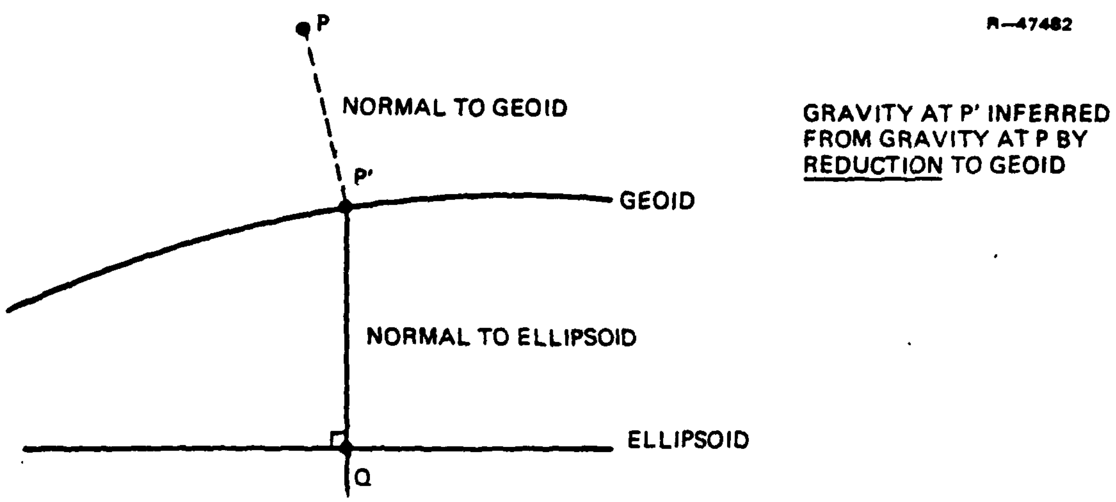
Figure 1.3-5 Gravity Anomaly for Point on Geoid

Local modeling of the gravity field then consists of a representation of measured gravity anomalies and deflection of the vertical values at a number of points, with some provision for interpolation to permit prediction of values at intermediate

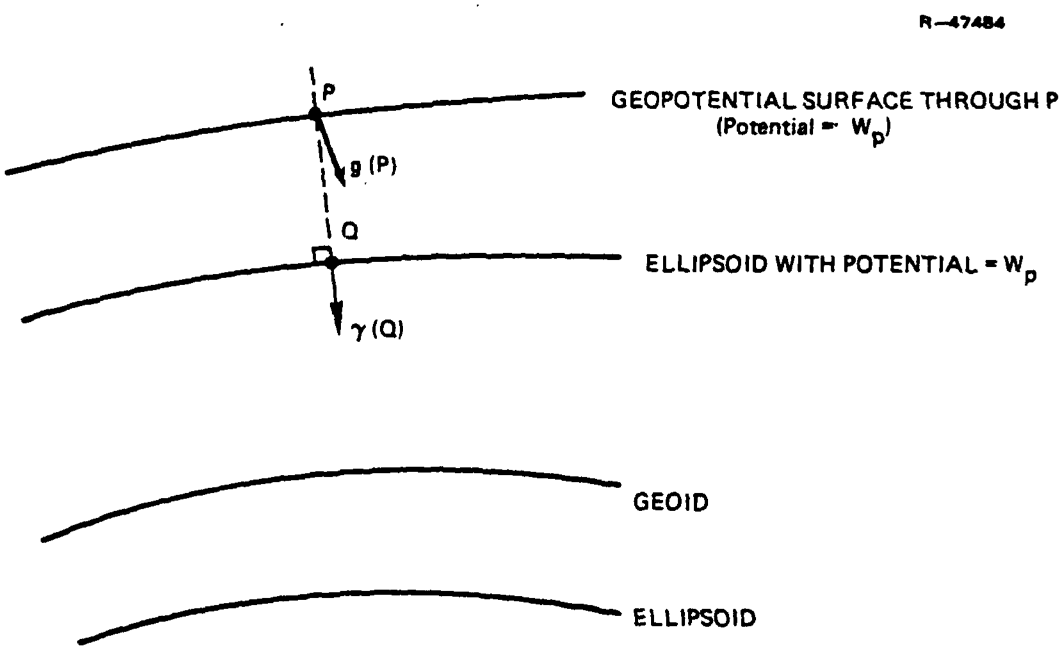
points of interest. A standard way of doing this has traditionally been by use of maps on which contours of equal gravity anomaly, inferred from the measured points, have been drawn. Similar maps are prepared for the components of the deflection of the vertical. More recently, digital data bases stored in computers have been used for this purpose. The gravity anomaly and deflection of the vertical results, however obtained, are combined with normal gravity for the point in question to give the gravity vector at that point. Gravity is computed from the normal gravity formula for the ellipsoid normal. How accurately this can be done depends primarily on the level of detail that be included in the gravity anomaly and deflection maps (or data bases) and is thus strongly influenced by practical considerations of time, cost, and data availability.

Because these concepts play an important role in discussions to follow later, it is necessary to give more precise and formal definitions of gravity anomaly, gravity disturbance, and deflection of the vertical. The gravity anomaly is defined first. For a point, P, on the geoid (at sea level), Fig. 1.3-5 shows a related point, Q, on the ellipsoid. The vector difference between the actual gravity at P, and the normal gravity at Q, is the gravity anomaly vector. The difference in direction between the two vectors is the deflection of the vertical. The scalar difference between the magnitude of the actual gravity at P and the normal gravity at Q is the gravity anomaly.

The definition of gravity anomaly is more complicated for a point not on the geoid (above the earth). The measured gravity at P must be reduced to the geoid by the use of formulas that describe the variation of gravity with height, taking into account the effect of mass (if any) between the geoid and



(a) REDUCTION TO GEOID



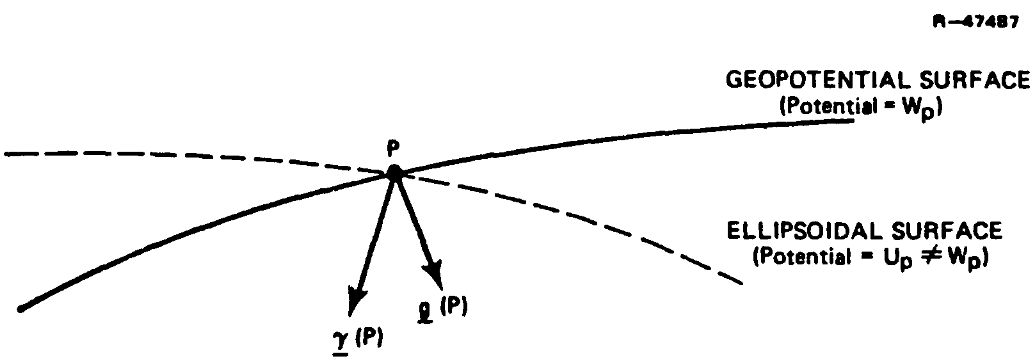
(b) EXTENSION OF GEOPOTENTIAL SURFACE TO P

Figure 1.3-6 Gravity Anomaly for Point Above Geoid

the point P. This is shown in Fig. 1.3-6A. A second approach* is shown in Fig. 1.3-6B. There is a surface of constant potential (in a sense, parallel to the geoid) passing through the point P on which the potential has a value W_P. An ellipsoid is constructed which requires the gravity potential on the ellipsoid to be equal to W_P. Then the anomaly and deflection of the vertical are defined as before in terms of the points P and Q.

The <u>gravity disturbance vector</u> is defined as shown in Fig. 1.3-7. The two surfaces involved are:

- A surface of constant geopotential passing through the point *P*
- An ellipsoidal surface also passing through the point P.



$\underline{\delta g} = \underline{g}(P) - \underline{\gamma}(P)$: GRAVITY DISTURBANCE VECTOR

$\delta g = g(P) - \gamma(P)$: GRAVITY DISTURBANCE

Figure 1.3-7 Gravity Disturbance Vector

*Although these two approaches may yield different answers in theory, such differences are generally negligible in practice.

Although the gravity disturbance may appear to be a simpler concept, the gravity anomaly traditionally has been more important in practice because it can be obtained directly: $\underline{g}(P)$ is either measured on the geoid or reduced to the geoid; $\underline{\gamma}(Q)$ is computed from the normal gravity formula for the ellipsoid.

1.3.3 Gravimetry and Gradiometry

Introduction - The gravity field of the earth may be represented by a scalar geopotential field, by the gradient of the geopotential (gravity vector), or by derivatives of higher order. Second derivatives of the geopotential can also be expressed as gradients of each component of the gravity vector. These gradients define the gravity gradient tensor. The relations are illustrated in Table 1.3-4.

TABLE 1.3-4
DIFFERENT CHARACTERISTICS OF THE EARTH'S GRAVITY FIELD

Gravitational Potential (V)
Gravitational Vector ($\underline{g}$) = $\underline{\nabla}V$
Gradient Tensor (Γ) = $\underline{\nabla}\ \underline{g} = \underline{\nabla}\ (\underline{\nabla}V)$
$\underline{\nabla}$ = gradient operator
$= \underline{i}\,\frac{\partial}{\partial x} + \underline{j}\,\frac{\partial}{\partial y} + \underline{k}\,\frac{\partial}{\partial z}$ (in rectangular coordinates)

Note: $\underline{i}$, $\underline{j}$, and $\underline{k}$ are unit vectors along the coordinate axes.

Direct measurements of the geopotential are usually accomplished by the satellite techniques discussed in the previous section and in later chapters. Measurements involving the magnitude of the gravity vector are usually referred to as

gravimetry. It can be shown that an error-free, continuous set of such measurements, taken over the entire surface of the earth, is sufficient (in principle) to determine the earth's gravity field everywhere on or above the surface.

Gravity Gradients - Gravity gradiometry is the term used for measurement of one or more elements of the gravity gradient tensor. Although the entire geopotential field could, in principle, be specified from worldwide measurements of one element of the gravity gradient tensor, the local character of gravity gradients precludes such an approach. Modern gradiometry generally involves measurement of all five independent elements of the gradient tensor.

Although there are nine elements of the gradient tensor, only five of these are independent. That is, the gravity gradient field can be completely specified at a spatial point by five appropriate elements of the gradient tensor. In the following paragraph this fact is illustrated by the use of rectangular coordinates.

The gravity gradient tensor is symmetric because of the commutativity of mixed partial derivatives, i.e.

$$\frac{\partial^2 V}{\partial x \partial y} = \frac{\partial g_x}{\partial y} = \frac{\partial g_y}{\partial x} = \Gamma_{xy} = \Gamma_{yx} \qquad (1.3\text{-}18)$$

hence three of the nine tensor elements are immediately seen to be redundant. The remaining redundancy is noted by observing the sum, S, of the diagonal elements of the gradient tensor (i.e., the trace)

$$S = \Gamma_{xx} + \Gamma_{yy} + \Gamma_{zz} \qquad (1.3\text{-}19)$$

Rewriting (1.3-19) in terms of second derivatives of the geopotential gives

$$S = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} \tag{1.3-20}$$

which is recognized as Laplace's equation. Since gravity fields are conservative, Laplace's equation applies everywhere outside of the surface of the body which generates the field. Hence S = 0 and only two independent measurements of the diagonal elements of the gradient tensor are required.

Gravity gradient measurements are frequently expressed in units called Eötvös (pronounced ēt vōs) and symbolized by the letter E. <u>One Eötvös represents a gravity change of one mgal in a span of ten kilometers</u>. Other equivalents are listed in Table 1.3-5.

TABLE 1.3-5
ALTERNATIVE DEFINITIONS OF THE EÖTVÖS UNIT (E)

One Eötvös
= 0.1 mgal/km
= 0.1 μgal/m
= 10^{-3} μgal/cm
= 10^{-9} gal/cm

Note that since one gal equals one cm/sec^2 the final entry in Table 1.3-5 can be expressed as

$$1.0\ E = 10^{-9}\ \text{gal/cm} = 10^{-9}\ (\text{cm/sec}^2)/\text{cm} = 10^{-9}\ \text{sec}^{-2}$$

Thus, gravity gradients can be expressed in fundamental units which involve only time.

Gravimeters - Gravimeters fall into two broad categories. Absolute gravity meters measure the magnitude of the gravity vector with very great accuracy, typically to a few microgal or better. Such measurements are used where gravity values of high precision are required (for example, to determine a gravity base station value at a key survey location or to study time-varying properties of the gravity field at geophysically interesting sites). Because of their expense, fragility, and time-consuming operating characteristics, absolute gravity meters are usually not employed for widespread gravity surveys. Instead, most gravity surveys are conducted with relative measuring instruments, carefully leveled or stabilized on a self-leveling platform. Usually, the term gravimeter refers to this type of device. Such instruments have scale ranges of plus and minus several hundred mgal with zero corresponding to nominal calibrated value of g (for example, 980 gal). The accurate determination of the value of gravity for zero gravimeter output typically involves calibration at absolute gravity meter measurement sites. Carefully performed gravimeter measurements are repeatable to better than 0.1 mgal. Such accuracies are routinely achieved for modern instruments like the LaCoste and Romberg Model S and the Bell BGM-2 gravimeter.

Unfortunately, instrument error is not the key source of inaccuracy in gravimeter surveys. Because the surface of the earth cannot be defined in analytical terms, gravity data are more usefully related to the geoid or to an ellipsoidal reference surface which approximates the geoid as discussed in Section 1.2.1). The reduction or downward continuation of data to such a reference surface requires assumptions about the local field which are rarely satisfied in practice. As a

result, errors as large as several mgal can be introduced into reduced gravity data such as gravity anomalies. The investigation of methods for treating gravity data to mitigate the effect of reduction errors is a current area of active research.

Traditionally, gravimeter measurements have been made at fixed, land-based sites but the need for ocean data has motivated development of systems which operate in a dynamic environment. Because the gravimeter cannot distinguish between vehicle accelerations and gravity,* vehicle motion must be accounted for and compensated to the greatest possible extent. Heave motion is typically of sufficiently high frequency to be filtered away, but the Coriolis acceleration due to the vehicle's velocity must be compensated (the Eötvös correction). In addition, the gravimeter measurements must be corrected for any maneuver-related accelerations. Despite the increased complexity, shipboard gravimetry has enjoyed considerable success and sizable portions of the ocean have been mapped to accuracies approaching one mgal.

Attempts to apply the same moving-base gravimetry techniques to aircraft have been unsuccessful for several reasons. The vibration and maneuver motion of aircraft typically involves larger accelerations than seagoing vessels. The problem of compensating for the kinematic accelerations is complicated by high aircraft velocities -- the gravity signal is no longer spectrally distinct from the acceleration noise. Higher aircraft velocities (and velocity errors) result in deterioration of the accuracy of the Eötvös correction as well. Finally the gravity signal itself is attenuated at high altitudes. At 20,000 ft (approximately 6 km), for example, the root

---

*Note that the same problem occurs in inertial systems -- see Section 1.2.7.

mean square value of the vertical gravity disturbance is about 80 percent of the surface root mean square value. These factors in combination have resulted in aircraft gravimeter survey errors in excess of 10 mgal, an accuracy level generally considered to be unacceptable. The need for quick and accurate mapping of the gravity field over large areas of the earth has encouraged examination of alternative ways to conduct airborne or satellite surveys. This has been one motivation for the development of gravity gradiometers.

Gradiometers - Devices which measure elements of the gravity gradient tensor, while available since the turn of the century, are currently not widely used. Although gravity gradients provide very detailed information about the local gravity field, the bulkiness, sensitivity, and immobility of the instruments, as well as time-consuming site preparation procedures, have resulted in relatively little application of gradiometers for practical field work. More recently, however, the advantages of measuring gravity gradients from a moving vehicle have prompted development programs for such instruments. Moving-base gradiometers presently under development, when used with an inertial navigation (or survey) system, will be able to distinguish between gravity and kinematic accelerations. If development is successful, gradiometers will measure local variations in the gravity field over large regions and conduct rapid, accurate gravity field surveys.

Although considerable progress has been made toward operational moving-base gradiometers, a viable sensor has not yet emerged from the laboratory. Currently, prototype gradiometers are under development by the Charles Stark Draper Laboratory and the Bell Aerospace Division of Textron, Inc. The Draper device measures two cross-gradient elements of the gradient tensor. Three instruments, aligned symmetrically and inclined

from the vertical by 35 degrees, provide enough linearly independent measurements to determine all of the gradient tensor elements. (Recall that only five elements of the tensor are required.)

The Bell gradiometer also provides two measurements, a cross-gradient and the difference between two of the diagonal axis gradients. Three Bell instruments oriented at right angles are required to measure the five independent elements of the gravity gradient tensor.

Despite the slow emergence of this technology, laboratory instruments have demonstrated noise levels near one E* and have been successfully operated under some simulated moving vehicle conditions. It is generally believed that, if successful, moving vehicle gradiometer technology will become operational during the mid 1980s.

*Such measurements usually involve averaging the noise over a specified period. Ten seconds is a benchmark often used for gradiometers.

CHAPTER FOUR
SATELLITE GEODESY

The subject of satellite geodesy deals with the use of artificial earth satellites to provide information about the questions with which the subject of geodesy is concerned:

- The size and shape of the earth
- The relative location of points on the earth's surface
- The detailed nature of the earth's gravity field.

Important geodetic and geophysical information has been obtained from a very large number of satellites -- beginning with the first Sputnik, launched on October 4, 1957 -- including a small number of geodetic satellites, some examples of which are:

- ANNA-1B, launched in 1962 (the acronym represents Army Navy NASA Air Force)
- The GEOS series (Geodetic Earth Orbiting Satellite), including GEOS-1 (1965), GEOS-2 (1968), and GEOS-3 (1975)
- LAGEOS (Laser Geodynamic Satellite), launched in 1976
- SEASAT* (1978).

The material covered in this Section includes:

- The two principal approaches to the use of artificial satellites for geodetic purposes -- termed geometric and dynamic

*SEASAT's primary mission was as an oceanographic, rather than geodetic, satellite. However, it has been used for geodetic purposes.

- The instrumentation and types of observational data used for satellite geodesy

- A geodetic application of laser ranging to the earth's natural satellite, the moon

- Results obtained from satellite geodesy; prospects for the future.

1.4.1 Methods of Satellite Geodesy

1.4.1.1 Geometric Satellite Geodesy

The geometric method uses the satellite (or satellites) as a target for surveying, and does not differ in principle from techniques for ground surveying. Geometric approaches are principally concerned with determination of the relative coordinates of tracking stations. They normally involve intervisible tracking of several passes of a satellite by a network involving known and unknown station locations.

The basic factor tying the network together is the simultaneous observations of points in space (satellites) from two or more stations. Figure 1.4-1 illustrates a possible geometric configuration. A and B are the known stations; C is the unknown. At least two sets of satellite positions (F_1 and F_2 in Fig. 1.4-1) must be observed simultaneously from all three stations. To provide better geometry, F_1 and F_2 should be from different satellite passes. The observed directions to the satellites from the three stations are combined with the known geodetic coordinates of A and B to obtain the rectangular coordinates of the satellite positions by celestial triangulation. The satellite positions then become known stations and with directions from C available, the position of C

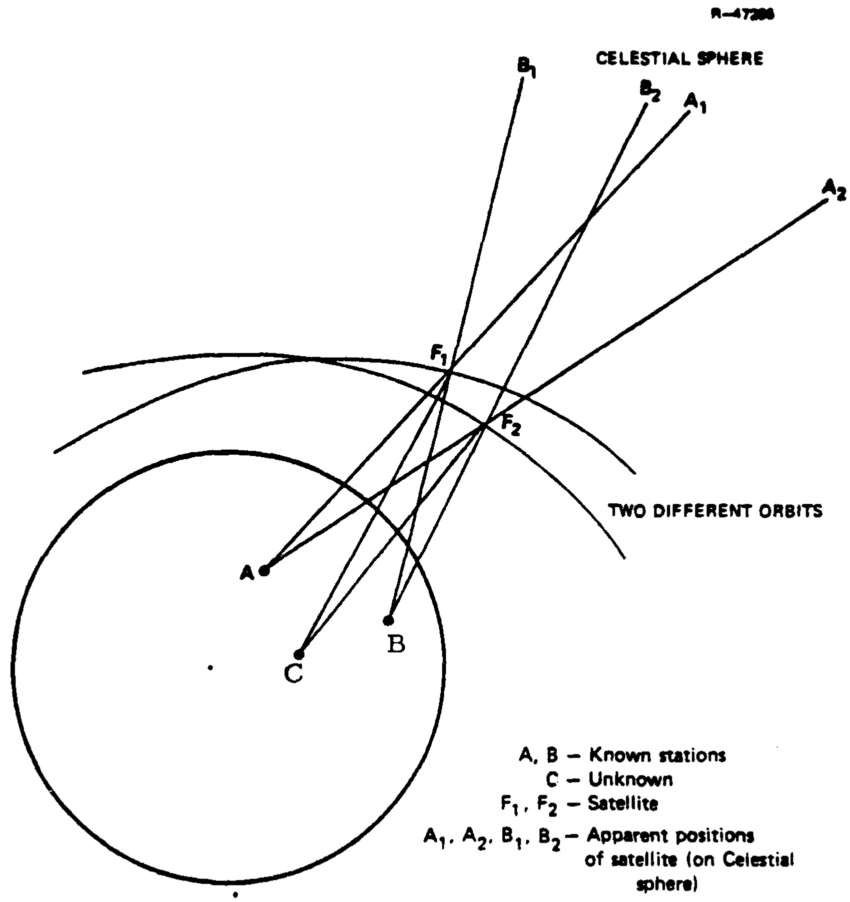


Figure 1.4-1 Simultaneous Observation Method With Angular Data

is computed. It then becomes part of the network to which A and B belong. Coordinates obtained by the geometric method are not geocentric. They depend on the position and precision of the known stations from which they were computed and, therefore, are subject to the same limitations as the coordinates of points in any geodetic system. In addition, errors in the measurement process will affect the accuracy of the unknown station.

If only angular measurements are available (as is usually the case with optical instruments), then one or more accurate baselines must be known to provide the scale for the survey. The lengths of the baselines need to be comparable to

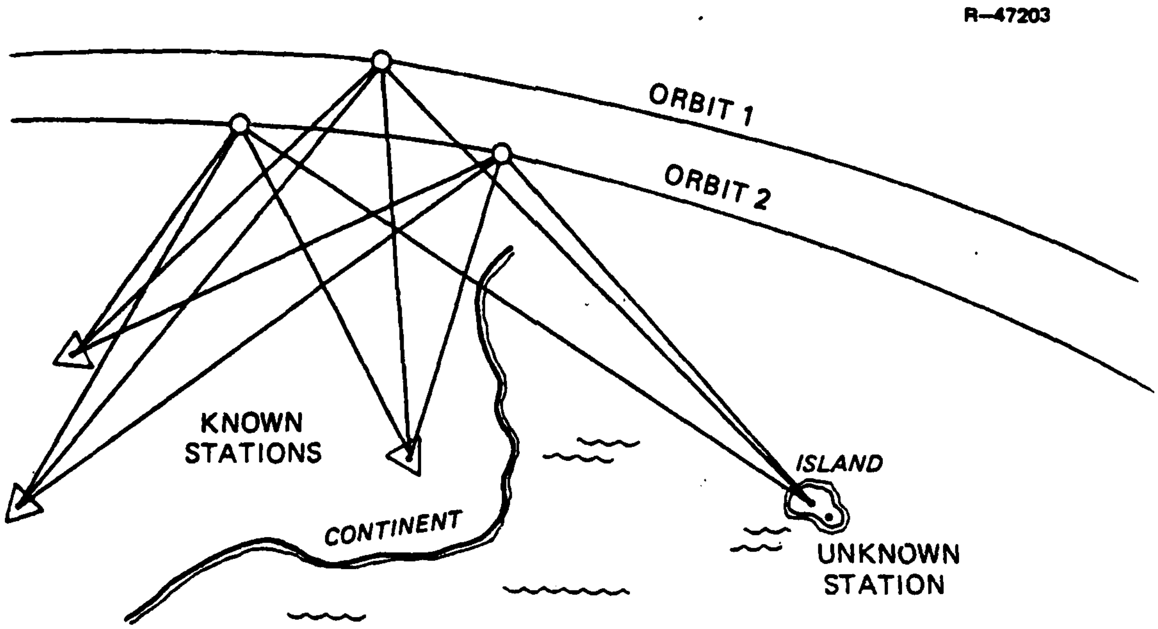


Figure 1.4-2   Simultaneous Observation with Range Data

the average distance between adjacent stations in the network. This average distance should approximate the altitude of the satellite observed.

If only range information is available as when satellites are observed using only radio signal delay information, then a configuration like that of Fig. 1.4-2 can be used. There are three known ground stations and one unknown station. The position of the satellite is fixed at the moment of observation by the simultaneous ranging information from the known stations. To insure good geometry, each satellite position observation should be established on a different pass of the satellite. The position of the unknown station is then computed from the simultaneously observed distances from it to the known satellite positions. Although two satellite passes are an absolute minimum, a large number of passes are usually

observed to secure the best geometric conditions. A least squares solution is then applied to the data to produce the best position for the unknown station. The coordinates of the station are in terms of the same datum as the known stations and, therefore, subject to the same limitations as those involved in the simultaneous solution from optical observations. Geodetic systems can be extended this way by a bootstrap approach. After enough observations are obtained for positioning one unknown point, it then becomes a known station and the mobile equipment at a former known site moves up to occupy the next unknown site.

1.4.1.2 Dynamic Satellite Geodesy

The dynamic method is based on the physical laws that govern satellite motion,* as well as the geometric configuration of satellites and tracking stations. In principle, the variation with time of any observed quantity (angle, range, range rate, etc.) associated with a satellite and an observing station can be modeled mathematically in the form

$$h = h\ (\alpha_i,\ \beta_i,\ \gamma_i,\ \delta_i,\ \varepsilon_i;\ t) \qquad (1.4\text{-}1)$$

where the observable, h, is given as a function of time and a large number of parameters grouped in the following categories:

- α_i are parameters defining the location of the observing station in the coordinate system used to model the motion of the satellite -- generally the stellar inertial coordinate frame
- β_i are parameters associated with the measuring instrument (bias or offset quantities, for example)

*Further material on the motion of earth satellites is included in Unit Four.

- γ_i are parameters defining the orbit of the satellite -- for example, the six classi[illegible]l or Keplerian elements
- δ_i are the geopotential coefficients defining the earth's gravitational field
- ε_i are other physical parameters to be determined or improved (for example, the velocity of light).

On the basis of observational data from many stations, involving many passes of a large number of satellites, values for the unknown parameters that lead to the best agreement between model and data are determined by a variety of computational techniques.

For best results with the dynamic method, satellites must be tracked periodically over a relatively long period of time with many successive observations. The known stations (tracking stations) should be well positioned on a worldwide basis and their coordinates periodically corrected with data derived from the computations. With such a network, unknown stations can be added and tied into the network after a few days of observations. It is also important to use satellites having a variety of orbital configurations for the purpose of determining the earth's gravitational field. Especially important is a wide range of orbital inclinations. Once the gravity field is known to sufficient accuracy, station positions can be determined from observations on a single satellite.

There are several variations of the dynamic method, characterized by the determination of only certain subsets of the parameters. For example, if only station positions are desired, the <u>semi-dynamic long arc</u> method may be used. This

requires a precise determination of the position of each satellite as a function of time. Such position time histories are often referred to as ephemerides. When the orbit ephemeris is available, then the data gathered during the observation period need not be used to compute the path of the satellite. The satellite track in space is considered to be without error and is held fixed during the solution for survey system coordinates. Stations positioned using this method are referred to the coordinate system in which satellite motion is computed, which is earth-centered to within the degree of accuracy stated by the source of the orbit determination. Since satellite orbits cannot actually be determined without error, the assumption is made that if enough passes (30 or more) are seen by the station in all directions and at all elevation angles, then the effects of the orbit errors will largely cancel out.

In another important modification of the dynamic method, called the short-arc method (or sometimes the semi-dynamic short-arc method), station coordinates and orbital parameters are solved for, but the coefficients of some previously determined gravitational model are accepted as known. The term short-arc stems from the fact that only portions of satellite arcs are observed -- the arc lengths usually being less than a quarter of a revolution. These arcs become the baselines for determining positions of the observers. In a short-arc solution, points along an arc are computed for the times of the observations as an intermediate step towards deriving the station positions. The values computed for these points are influenced by the gravitational model, the observations from the ground stations, and an initial state vector composed of three position and three velocity coordinates. The coordinates refer to a point usually chosen near the center of the arc. Since the state vectors* do not have to be

*That is, the sets of position and velocity coordinates.

known exactly (they are determined in the course of the solution), an observer is free to use other satellites besides those for which precise orbits are available. This allows for a more economical survey project as the needed observations can be acquired over a shorter time period.

If appropriate constraints are not placed on the initial state vectors, then additional constraints are needed (involving at least three stations) to define the coordinate system of the survey. For example, the positions of three stations can be held fixed, or the position of one station can be fixed in conjunction with constraints on some orientation parameters such as station-to-station azimuth and elevation angles.

If effective constraints are known and are placed on the state vectors, then no constraints need be applied to the coordinates of the stations. At least one station should be continuously occupied during a survey project, while other receivers can move from station to station. The purpose of this scheme is to effect the interrelation of all stations through connections to the permanent station or stations.

The short-arc method is used to best advantage when two or more stations observe passes in common. Four well-distributed stations observing the same pass can yield adjusted orbits of high accuracy. A single-station short-arc reduction, in which the orbital state vectors are held fixed,* would correspond to the semi-dynamic long-arc method.

*That is, satellite position and velocity time histories are assumed to be known, and are not solved for in the course of defining the station coordinates.

1.4.2 Observational Data for Satellite Geodesy

Some of the more important systems used to obtain satellite-related measurements for use in geodesy are now reviewed. In general, these systems serve to measure angle, range, or range-rate information.

Optical tracking consists of photographic observations of a satellite against a background of stars. Typical instrumentation consists of special telescopic cameras such as the BC-4, PC-1000, MOTS, or Baker-Nunn. Large satellites with no independent source of illumination can be photographed after dusk or before dawn when sunlight illuminates the vehicle against a dark sky. Other satellites are equipped with flashing lights whose firing times are carefully controlled. Since star positions are recorded to a high degree of precision in star catalogs, the background stars -- once identified -- provide a framework on the photographic plate or film for a determination of precise directions from camera station to satellite. Analytical photogrammetric methods (refer to Section 1.2.3) are used to derive the direction. Camera observations are used in both geometric and dynamic methods. In either case, the timing record is a vital part of the data obtained. The clock time must be carefully controlled at every station to achieve accurate results.

Range tracking is illustrated by the SECOR system (Sequential Collation of Range), developed by the U.S. Army. The first SECOR transponder was orbited on ANNA-1B in 1962. The SECOR system continued in use through 1970. It consists of four ground stations and an earth orbiting satellite. The system operates on the principle that an electromagnetic wave propagating through space undergoes a phase shift proportional to the distance traveled. A phase modulated signal transmitted

from a ground station is received by the satellite-borne transponder and then returned to the ground. The phase shift experienced by the signal during the round trip from ground to satellite and back to ground is measured electronically at the ground station. From the phase delay a digitized representation of range is provided. Other examples of range tracking include radar installations of various kinds, operating in conjunction with beacons (transponders) on the satellite, or skin-tracking passive or uncooperative satellites.*

Range-rate tracking is often called Doppler tracking, because it depends on the Doppler effect. While a satellite transmitter sends a continuous unmodulated wave at a fixed frequency, the received signal at the tracking stations exhibits a shift in frequency due to the relative velocity of the satellite and observing station. A similar phenomenon may be observed with sound waves, as the source of the sound approaches and recedes from the observer. Although the sound waves travel at a constant rate, they become crowded together as the source approaches the observer, the wave lengths become shorter, and the pitch increases. The opposite effect takes place as the source moves away. The frequency change is described by the equation

$$\frac{\Delta f}{f} = \frac{v}{c} \tag{1.4-2}$$

where

Δf = frequency change

f = original frequency

---

*Skin-tracking refers to tracking based on the reflected radar beam. Beacon tracking refers to tracking performed with a cooperative vehicle which transmits a return signal to the ground station upon receiving a signal from the ground.

v = relative source-observer velocity

c = velocity of signal propagation (light or sound)

The frequency received at the tracking station is a function of the transmitted frequency, velocity of propagation, and the rate of change of the slant range (i.e., straight line distance) between satellite and station. From observations at one station, the satellite's period, time of closest approach, distance of closest approach, and relative velocity with respect to the station can be determined. If observations are made from three or more known stations, the orbital parameters may be derived.

In practice, receiving equipment does not measure instantaneous range-rate or Doppler. Rather, the receivers count carrier cycles received over a given time interval (indirectly, after mixing the received signal with a ground-generated signal to decrease the frequency). Although the ratio of the cycle count to the interval is an approximation of the frequency, the duration of the count is too long to allow neglect of nonlinearity in the frequency. Therefore, the cycle counts are either directly used as observational data, or, more often, converted to range differences (rather than range rates).

Doppler tracking has been a very fruitful source of data for satellite geodesy, for a number of reasons:

- It is passive, requiring neither an interrogation nor directionally sensitive antennas at the receiver
- The data obtained (Doppler counts) are in digital form

- The radio frequencies used permit all-weather day and night tracking
- Accuracies have steadily improved.

The Defense Mapping Agency operates a worldwide Doppler tracking network called TRANET with stations at several permanent sites. Automatic portable receiving equipment for Doppler tracking is available from several suppliers. For example, the GEOCEIVER (geodetic receiver) has played an important role in various survey projects. Positions of thousands of sites determined with the use of Doppler receivers have provided a global geodetic network. The Defense Mapping Agency has used Doppler data routinely for the determination of polar motion since 1970. Since the Bureau International de L'Heure considers these data to be the most accurate currently available, its final results are heavily weighted to the Doppler data.

Laser ranging systems measure the time interval between an outgoing pulse and the reflection of the pulse from a satellite. The time interval is measured very accurately and then transformed into a range measurement that is corrected for atmospheric refraction. Electronic and mathematical corrections are applied to ensure that the time interval measurement is taken between identical portions of the outgoing and returning pulses. Laser ranging is possible even when the satellite is in the earth's shadow as well as during daylight hours. The satellites used as targets by laser system are usually equipped with special retroreflectors (corner cube reflectors*) to enhance reflectivity.

---

*Corner reflectors are described in more detail in Section 1.4.3.

Simultaneous laser ranging to a near-earth satellite from two sites is used to determine the coordinates of one laser site relative to the fixed position of the other site. The intersite distance is also determined. The National Aeronautics and Space Administration (NASA) has used laser tracking since 1972 to measure the distance between points in North America. One application includes tests to determine the accuracy of laser tracking in measuring the crustal motion between points on opposite sides of the San Andreas fault. Repeated measurement of baselines across the fault are involved over a period of several years. Simultaneous laser tracking has also been achieved between United States east coast sites and Bermuda. Recall from Section 1.2.2 that this makes possible a determination of the Bermuda site's relative location and the baseline between Bermuda and each coastal site. Laser ranging data have been incorporated into the development of world geodetic systems by the Smithsonian Astrophysical Observatory (SAO) and the Department of Defense (DoD). NASA has also included laser data in their development of gravitational models.

Satellite-to-satellite tracking is a relatively new concept with a variety of possible implementations. A high-altitude satellite (in synchronous or near-synchronous orbit) may act as a relay from ground tracking stations to a low-altitude satellite. In this way, low-altitude satellites may be observed when they are not accessible to ground stations. With this type of tracking, the signal generated by a tracking station is received at the relay satellite and then retransmitted to the lower altitude satellite. A return signal is then sent from the low satellite back to the high satellite and on to the ground station. In another variation, two low-altitude satellites track one another, observing mutual orbital variations caused by gravity field irregularities. Another

concept calls for several high altitude satellites with accurately known orbits to fix the position of a low-altitude satellite. More detail is provided in Unit Four.

Satellite-to-satellite tracking data are currently being collected and analyzed in a high-low configuration between the ATS-6 radio relay satellite and GEOS-3, a low altitude geodetic satellite. This system is being studied to evaluate its potential for both orbit and gravitational model refinement.

<u>Satellite radar altimetry</u> is used for direct measurement of the geoid in ocean areas. The satellite altimeter consists of a downward ranging radar that measures the time delay from the transmission to the reception of a pulse of energy. The apparent one-way distance from the transmitting antenna to the surface is equal to one-half the product of the time delay and the speed of light. From this distance or height, the local surface effects such as tides, winds, and currents are removed to obtain the satellite height above the geoid. Precise knowledge of satellite position from other tracking data sources permits the determination of the distance from the center of the earth

- To the satellite
- To the ellipsoid

along the same line on which the satellite's height above the geoid was obtained. Thus the geoid can be related to the reference ellipsoid, giving the <u>geoid height</u>, or <u>undulation</u>, a quantity of direct geodetic interest.

The SKYLAB spacecraft, launched in 1973, was the first attempt at satellite-based radar altimetry. This was a research

mission from which information was obtained for the design of future altimeter instruments. The GEOS-3 altimeter incorporated many of the design features that were tested in SKYLAB. Launched in 1975, GEOS-3 provided geoid measurements over the water areas of the earth from 65 deg N to 65 deg S. A considerably improved altimeter, launched in the SEASAT satellite in June 1978, returned voluminous data of extremely high quality until the premature failure of the spacecraft's power system.

1.4.3 Lunar Laser Ranging

Introduction - The concept of lunar laser ranging involves a precise measurement of the distance between a point on the surface of the earth and a point on the surface of the moon. This measurement is made by sending a laser beam (a pulse of light) from the surface of the earth to a specially designed reflector placed on the moon during one of the lunar missions and by measuring the time interval required for the laser beam to complete its round-trip. The round-trip time is approximately 2.4-2.7 seconds. By multiplying the time interval by the speed of light and dividing by 2 to account for the round trip, the distance between the two points can be measured. Since the time intervals can be measured to an accuracy of approximately 1 nsec, the corresponding range measurement accuracy is about 15 cm.

Laser ranging measurements of this accuracy have been important to scientists for several applications. The principal contributions of lunar ranging to geodesy, thus far, have been to provide a more accurate estimate of the parameter GM* which is the product of the gravitational constant, G, and the

*See Section 1.3.2.

mass of the earth, M, and to provide more precise data concerning the rotation of the earth. Other important contributions in lunar astronomy have been to enable scientists to determine a more precise description of the orbit of the moon about the earth and of the angular rotation of the moon about its axis.

Background - The concept of lunar laser ranging was proposed first in 1962 for use on one of the early Ranger missions. Ranger was a NASA program of unmanned spacecraft launched on trajectories designed to impact on the moon. The Ranger spacecraft provided the first close-up photography of the moon which was later used in planning the Apollo lunar missions. The first lunar laser reflector was placed on the moon in a "semi-soft" Ranger impact. Successful processing of these first lunar laser signal returns was reported in 1962. This is of historical interest only. The quality of later laser systems made the above data obsolete.

The next lunar laser reflector system was placed on the moon by astronauts Neil Armstrong and Edwin Aldrin during the Apollo 11 mission. Additional laser reflectors were placed on the moon by both American and Russian Luna spacecraft. Their positions and the mission which placed them are indicated in Fig. 1.4-3.

The availability of several reflectors on the moon provides capability for scientists to measure with high precision relative changes in the distances from their ground stations to these lunar locations. This information provides them with data which can be used to determine changes in both the lunar rotation and the earth's rotation.

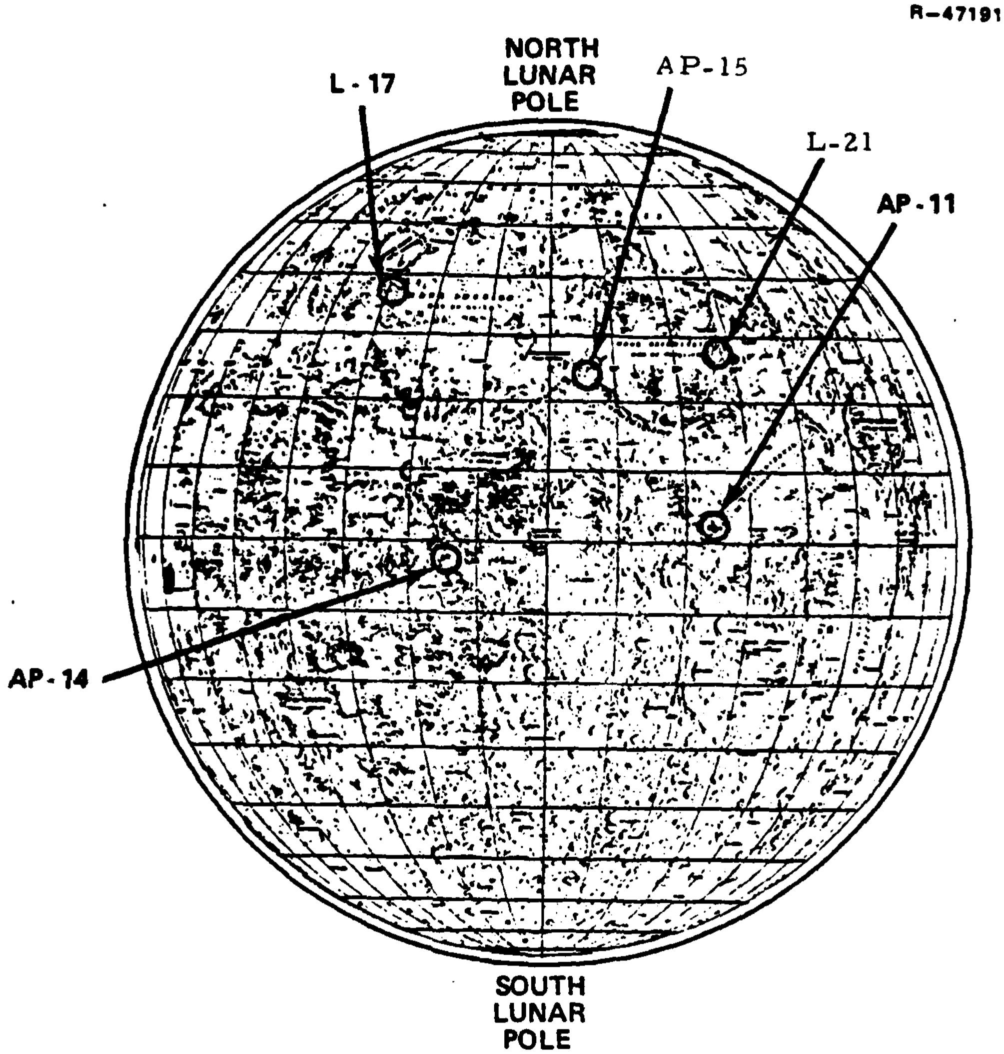


Figure 1.4-3 Locations of Lunar Laser Reflectors (AP denotes Apollo, L denotes Luna)

Lunar Laser Reflectors - A lunar laser reflector is a simple device. It is designed as the corner of a cube. When the laser beam strikes the surface of the reflector, it is reflected in precisely the same direction from which the beam originated. This helps to ensure that the antenna that was used to transmit the beam is also able to receive it.

The surface of the laser reflectors is made of highly polished fused silica. This surface is designed specially to provide high reflectivity for the laser beam.

A schematic diagram of a lunar laser reflector is shown in Fig. 1.4-4. This figure illustrates the reflection of the beam from its surface.

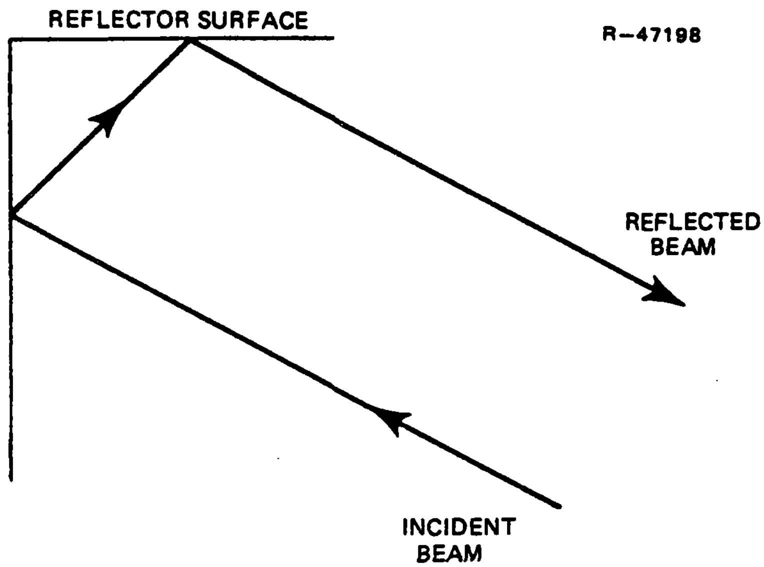


Figure 1.4-4 Schematic of Lunar Laser Reflector

Each of the lunar laser reflecting systems is composed of an array of these corner cube reflectors. The largest reflector was placed on the moon by the Apollo 15 astronauts. Its dimensions are 104 cm by 61 cm, and it contains 300 corner cubes.

<u>Lunar Laser Ranging Operations</u> - Lunar laser ranging data have been collected by many organizations. Scientists at the MacDonald Observatory of the University of Texas, the Smithsonian Astrophysical Observatory in Cambridge, MA, the Air Force Geophysics Laboratory at Hanscom Air Force Base, MA, the Ecole Polytechnique in Paris, France, and in the Soviet Union have reported results from the analysis of laser ranging data.

Observing periods can be scheduled for approximately three weeks per month. Nights of the month close to a new moon are avoided because of poor antenna pointing control. Other limitations on operating periods may result from periods of clouds and rain.

During a successful operation, the laser transmits pulses at time intervals of approximately 2-3 seconds. Since the round-trip time is approximately 2.4-2.7 seconds, there is no ambiguity in determining which return pulse is connected to which transmitted pulse. The accuracy of the timing circuits which measure round-trip time has been improved during the operation of the laser ranging systems and can now be calibrated to within 0.1 nsec, with an overall system accuracy of approximately 1 nsec. The laser signal is a high energy pulse with a width of approximately 2-4 nsec and a single-pulse output energy of approximately 3 joules.

Principal Results - As stated in the Introduction, the principal results for geodesy in the lunar laser ranging program have been in the determination of GM and in the measurement of the earth's rotation. At this writing, more than 9 years of lunar laser ranging data have been collected. Other techniques such as VLBI, discussed in an earlier section, are also being used to provide related measurements. The results obtained have allowed scientists to get progressively more accurate measurements. For example, a current estimate of the parameter GM derived from lunar laser ranging data is given by

$$GM = 398600.46 \pm 0.03 \ km^3/sec^2$$

Current results from the analysis of laser ranging data have enabled scientists to determine the earth's angular position in space to an accuracy of 0.01 $\widehat{\text{sec}}$ (corresponding to

7×10^{-4} sec of time). This gives an agreement with astronomical techniques to approximately 2×10^{-3} $\widehat{sec}$ over an average period of 5 days. The data have shown that there is little evidence of rapid variations in the earth's rotation. Significant further advances are expected from increasing the accuracy of determination of the earth's rotation and the variation in the universal time scales as more accurate lunar laser ranging data are collected over long periods of time.

Prospects for the Future - New lunar laser ranging stations are being established in Hawaii, Australia, Japan, France, and West Germany. These newer stations include plans for the development of lasers with 200 picosec (1 picosec = 10^{-12} sec) pulses and 3 cm accuracy. Furthermore, a broader geographical distribution of these stations will lead to increased accuracy in the determination of GM and the earth's rotation. As scientists continue to collect lunar laser ranging data, their understanding of variations in the earth's rotation and gravity field, as well as the moon's orbit, rotation, and mass, will continue to improve. The lunar laser ranging system is one of the many significant scientific results from the lunar landing programs.

1.4.4 Results of Satellite Geodesy

The use of artificial satellite data during the two decades that have passed since the first launches has led to an enormous increase in knowledge of the size and shape of the earth and the detailed structure of the gravity field. Some indication of the level of detail with which the gravity field can be modeled at any given time is given by the number of geopotential coefficients* known. A few months of tracking

*Review Section 1.3.2.

the first artificial satellite (in 1957) led to greatly improved knowledge of the C_{20} term (and also the semi-major axis and flattening of the reference ellipsoid). Not long thereafter several additional zonal harmonic terms were computed, and by 1962 about eight zonal terms and a few tesseral terms were published.

A 1962 solution of this type developed by the Naval Surface Weapons Center (NSWC) was used in 1963 in satellite orbit computations required to position LORAN-C sites in the Pacific. In 1965, NSWC developed over 144 terms (complete to degree and order 12) in the gravity model; this model was an important contribution to the development of Department of Defense World Geodetic System 1966 (DoD WGS-66), and was used for several years in orbit computations used in developing positions of sites in a worldwide geodetic network to 10 m accuracy. These solutions were based primarily on Doppler data.

In 1966 an earth gravity model based mainly on optical satellite data was published by the Smithsonian Astrophysical Observatory (SAO). It included 123 geopotential coefficients (the model was complete to degree and order 8, with zonal terms to degree 13, and a few tesseral terms of degrees 9 through 15). Six years later, SAO's Standard Earth III, based on camera and laser data, and combining satellite and surface gravimetry results, included

- Zonal terms up to $C_{23,0}$ with $C_{35,0}$ and $C_{36,0}$ also determined
- Approximately 360 tesseral terms (complete through the 18,18 terms, with additional terms up to degree 24).

By 1970, NSWC expanded the gravity field to include nearly 500 terms; this solution (based primarily on Doppler data) formed a basis for DoD-WGS-72 and allowed positioning of sites to about 1 m accuracy. As a final example, an improved Goddard Earth Model (GEM 10B) was published in 1978, complete through degree and order 36. It was based on satellite tracking data, surface gravity measurements, and satellite altimeter measurements from GEOS-3.

With the increased use of new measurement types -- satellite altimetry, satellite-to-satellite tracking, and laser ranging -- satellite geodesy will continue its contribution to geodetic and geophysical knowledge in the future.

APPENDIX A.1

FIELDS AND POTENTIAL

The concepts of scalar field, vector field, and potential -- used throughout Unit One -- are reviewed briefly for the convenience of the reader.

Scalar and vector fields - The term *field* is used to describe a physically observable quantity that is defined at every point of some region of space. Examples of regions in which field quantities might be defined:

- All of space
- The interior of a sphere, including the surface
- The exterior of a sphere, excluding the surface.

A *scalar field* is defined by a single number at each point -- for example, temperature or humidity. A *vector field* has magnitude and direction at each point, and requires the specification of *three* numbers (the rectangular coordinates, for example). A scalar field over a region is specified, therefore, by a single function of the three space variables:

$$u = u(x,y,z) \tag{A.1-1}$$

While a vector field requires *three* functions of the space variables:

$$\underline{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \qquad (A.1\text{-}2)$$

$$v_x = v_x(x,y,z) \qquad (A.1\text{-}3)$$

$$v_y = v_y(x,y,z) \qquad (A.1\text{-}4)$$

$$v_z = v_z(x,y,z) \qquad (A.1\text{-}5)$$

The gravity field - The gravity field on or near the surface of the earth is a specific example of a vector field. The field quantity is the acceleration due to gravity, varying in a complicated way with position but generally decreasing with distance from the center of the earth.

In the case of a point mass, the associated gravity field is characterized by:

- A magnitude inversely proportional to the square of the distance from the mass point
- A direction pointing inward toward the mass point.

For a point of mass M placed at the origin of a rectangular coordinate system (Fig. A.1-1), the field of acceleration is described by the equations

$$a_x = - \frac{GM}{r^3} x \qquad (A.1\text{-}6)$$

$$a_y = - \frac{GM}{r^3} y \qquad (A.1\text{-}7)$$

$$a_z = - \frac{GM}{r^3} z \qquad (A.1\text{-}8)$$

R-50973

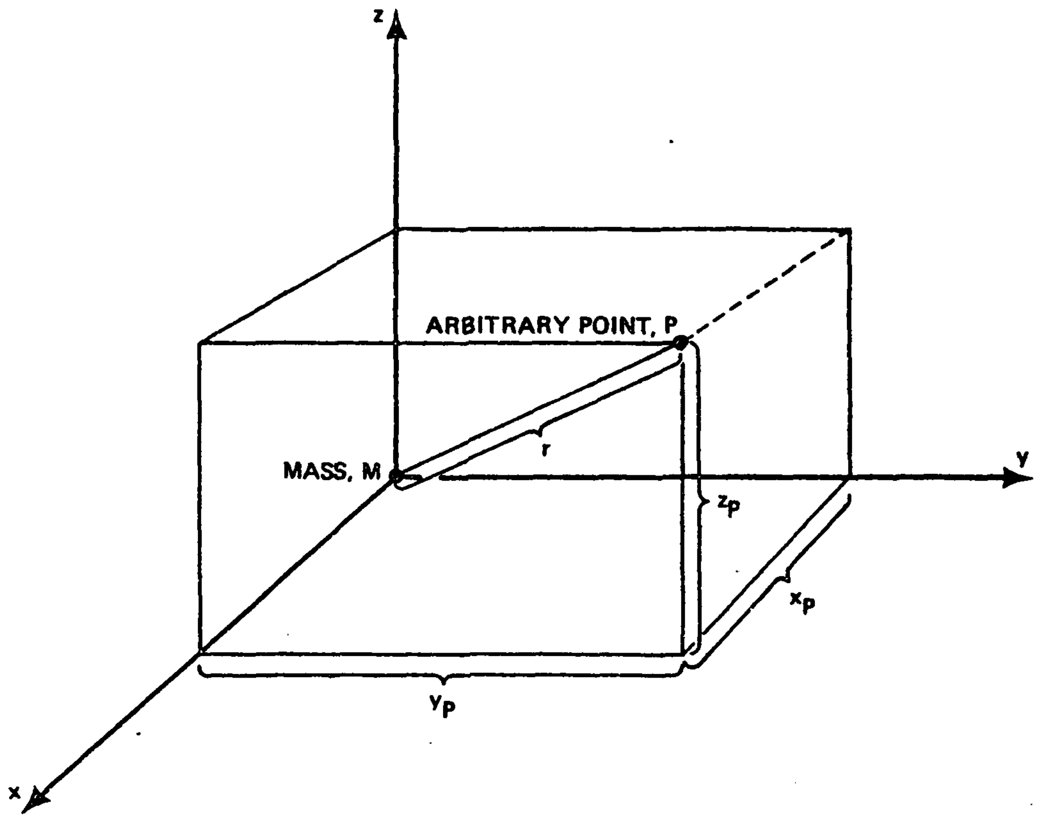


Figure A.1-1 Field of Acceleration Due to a Point Mass

where:

a_x, a_y, a_z = components of acceleration (m/sec^2)

x, y, z = coordinates of point at which acceleration is measured (m)

$r = (x^2 + y^2 + z^2)^{\frac{1}{2}}$

M = mass (kg)

G = universal constant of gravitation ($6.67 \times 10^{-11}\ m^3/kg \cdot sec^2$)

The acceleration due to a pair of point masses (see Fig. A.1-2) is the vector sum of the individual accelerations, each having the form of Eq. A.1-6 through A.1-8. For example, the x-component of the acceleration at point P due to the mass M_1 is

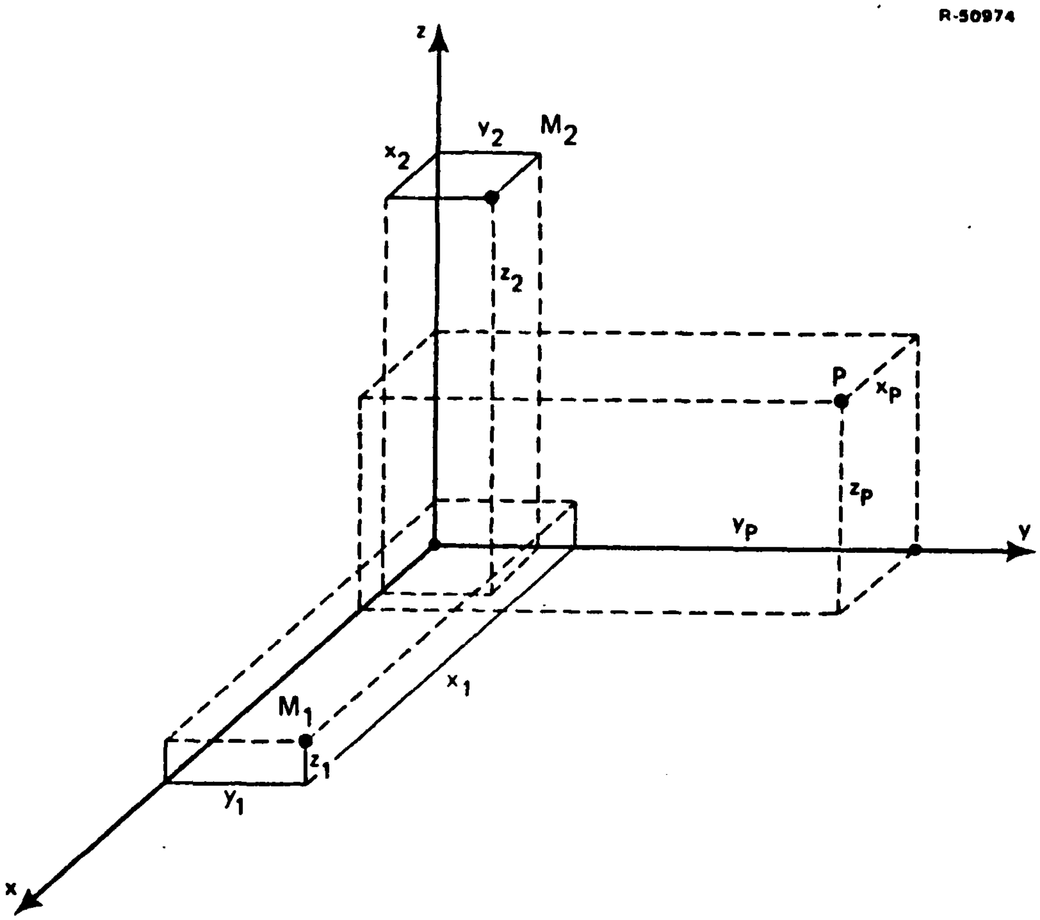


Figure A.1-2 Field of Acceleration Due to a System of Point Masses

$$(a_x)_1 = -\frac{GM_1}{r_{1p}^3}(x_p - x_1) \qquad \text{(A.1-9)}$$

and the x-component of the acceleration at point p due to the mass M_2 is

$$(a_x)_2 = -\frac{GM_2}{r_{2p}^3}(x_p - x_2) \qquad \text{(A.1-10)}$$

where:

$$r_{1p} = [(x_1-x_p)^2 + (y_1-y_p)^2 + (z_1-z_p)^2]^{\frac{1}{2}} \qquad \text{(A.1-11)}$$

and

$$r_{2p} = [(x_2-x_p)^2 + (y_2-y_p)^2 + (z_2-z_p)^2]^{\frac{1}{2}} \qquad \text{(A.1-12)}$$

The x-component of the resultant acceleration at point p is then given by the sum:

$$a_x = -G\left[M_1 \frac{x_p - x_1}{r_{1p}^3} + M_2 \frac{x_p - x_2}{r_{2p}^3}\right] \tag{A.1-13}$$

For a system of n point masses, Eq. A.1-13 generalizes to the form

$$a_x = -G \sum_{i=1}^{n} M_i \frac{x_p - x_i}{r_{ip}^3} \tag{A.1-14}$$

A solid body may be approximated by a large number of masses, each within a small volume element (Fig. A.1-3), and the acceleration field external to the body is represented by three sums (one for each component) having the form of Eq. A.1-14. By the use of limiting processes familiar from the calculus, the approximating sums become integrals, and Eq. A.1-14 takes on the limiting form

$$a_x = -G \iiint_V \frac{(x_p - x)\,\rho\,dx\,dy\,dz}{[(x_p - x)^2 + (y_p - y)^2 + (z_p - z)^2]^{3/2}} \tag{A.1-15}$$

where the integral extends over the volume occupied by the body. The notation is defined in Fig. A.1-3. Note that the mass of a volume element is given by

$$dm = \rho\,dx\,dy\,dz \tag{A.1-16}$$

where ρ is the density (measured in units of mass per unit volume).

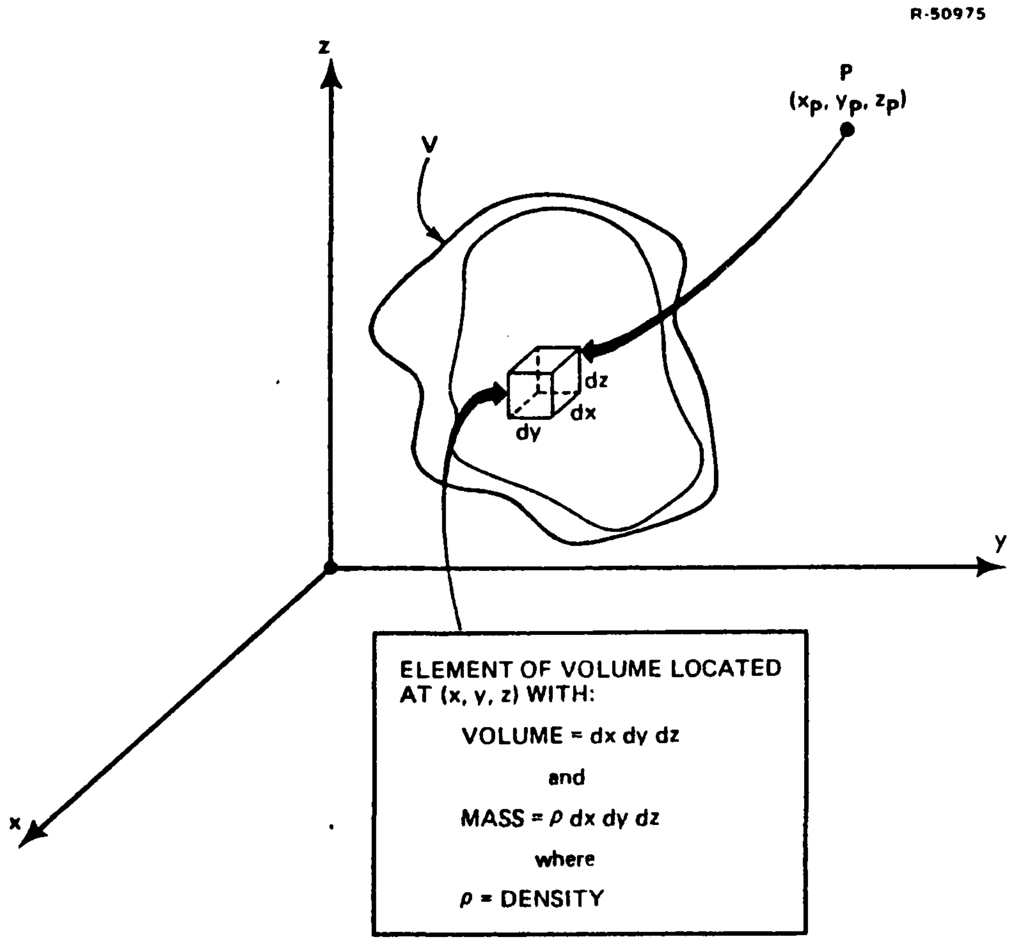


Figure A.1-3   Field of Acceleration Due to a Solid Body

If the body is a sphere, and if the density depends only on the distance from the center of the sphere,* then the acceleration at an external point is the same as if all the mass were concentrated at the center. This remarkable fact was proved by Isaac Newton in the 17th century.

Conservative fields and potential - The vector field characterizing the external gravitational acceleration due to

*This condition of radial symmetry is automatically satisfied, of course, if the density is constant (uniform) throughout the body.

any arbitrary body (or collection of point masses) has an important special property that greatly simplifies gravity calculations. Fields having this property are called conservative. Such a field can be described by a scalar quantity known as a potential. More specifically, the vector is equal to the gradient of the scalar potential:

$$\underline{a} = \text{grad } U = \begin{pmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial U}{\partial y} \\ \frac{\partial U}{\partial z} \end{pmatrix} \qquad \text{(A.1-17)}$$

where:

$\underline{a}$ = a conservative vector field quantity (for example, the acceleration due to gravity)

U = the associated scalar potential (for example, the gravity potential)

The significance of the conservative property is that a single function of the space variables (the potential) suffices for a complete characterization of a vector field that otherwise would require three functions for its specification.

Gravitational Potential - The acceleration vector associated with a point mass, as given in Eqs. A.1-6 through A.1-8, can be expressed in terms of a potential, U:

$$\underline{a} = \text{grad } U \qquad \text{(A.1-18)}$$

where

$$U = \frac{GM}{r} \qquad \text{(A.1-19)}$$

as the reader may verify by differentiation of Eq. A.1-19. Similarly, the acceleration due to a pair of point masses (Eq. A.1-13) is described by the potential

$$U = G\left(\frac{M_1}{r_{1p}} + \frac{M_2}{r_{2p}}\right) \tag{A.1-20}$$

and the acceleration due to n point masses (Eq. A.1-14) by the potential

$$U = G \sum_{i=1}^{n} \frac{M_i}{r_{ip}} \tag{A.1-21}$$

As a final example, the gravitational acceleration associated with a solid body (Eq. A.1-15) is derived from the potential

$$U = G \iiint_V \frac{\rho \; dx \; dy \; dz}{[(x_p-x)^2 + (y_p-y)^2 + (z_p-z)^2]^{\frac{1}{2}}} \tag{A.1-22}$$

APPENDIX B.1

LEGENDRE POLYNOMIALS AND ASSOCIATED LEGENDRE FUNCTIONS

The Legendre polynomials are defined as follows:

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2-1)^n \qquad \text{(B.1-1)}$$

where n is the degree of the polynomial. Some examples are:

$$P_0(x) = 1 \qquad \text{(B.1-2)}$$

$$P_1(x) = x \qquad \text{(B.1-3)}$$

$$P_2(x) = \frac{3}{2} x^2 - \frac{1}{2} \qquad \text{(B.1-4)}$$

$$P_3(x) = \frac{5}{2} x^3 - \frac{3}{2} x \qquad \text{(B.1-5)}$$

$$P_4(x) = \frac{35}{8} x^4 - \frac{15}{4} x^2 + \frac{3}{8} \qquad \text{(B.1-6)}$$

For computational purposes, recursion formulas like the following are used:

$$P_n(x) = -\frac{n-1}{n} P_{n-2}(x) + \frac{2n-1}{n} x P_{n-1}(x) \qquad \text{(B.1-7)}$$

starting with the simple expressions for P_0 and P_1.

The associated Legendre functions may be derived from the Legendre polynomials by the following formula:

$$P_{nm}(x) = (1-x^2)^{m/2} \frac{d^m P_n(x)}{dx^m} \qquad (B.1\text{-}8)$$

but are usually computed from recursion formulas analogous to Eq. (B.1-7). Since the associated Legendre functions appear in Eq. (1.3-12) with

$$x = \sin \phi' \qquad (B.1\text{-}9)$$

as the argument, examples will be given of $P_{nm}(\sin \phi')$ as a function of ϕ':

$$P_{11}(\sin \phi') = \cos \phi' \qquad (B.1\text{-}10)$$

$$P_{21}(\sin \phi') = 3 \cos \phi' \sin \phi' \qquad (B.1\text{-}11)$$

$$P_{22}(\sin \phi') = 3 \cos^2 \phi' \qquad (B.1\text{-}12)$$

$$P_{31}(\sin \phi') = \cos \phi' \left[\frac{15}{2} \sin^2\phi' - \frac{3}{2}\right] \qquad (B.1\text{-}13)$$

$$P_{32}(\sin \phi') = 15 \cos^2 \phi' \sin \phi' \qquad (B.1\text{-}14)$$

$$P_{33}(\sin \phi') = 15 \cos^3 \phi' \qquad (B.1\text{-}15)$$

For some purposes it is advantageous to modify the definition of the Legendre functions, through multiplication by a numerical constant, in order that the mean square value of any surface spherical harmonic, over the entire surface of the sphere, be equal to one. This is done by multiplying P_{nm}, as defined above, by the normalizing factor

$$K_{nm} = \left[2(2n+1) \frac{(n-m)!}{(n+m)!}\right]^{\frac{1}{2}} \qquad (B.1\text{-}16)$$

for $m \neq 0$, and

$$K_{n0} = (2n+1)^{\frac{1}{2}} \qquad (B.1\text{-}17)$$

for m=0. The resulting normalized Legendre functions

$$\bar{P}_{nm} = K_{nm} P_{nm} \tag{B.1-18}$$

may be used in Eq. (1.3-12) in place of the ordinary functions, provided that the C_{nm} and S_{nm} coefficients are correspondingly adjusted:

$$\bar{C}_{nm} = \frac{1}{K_{nm}} C_{nm} \tag{B.1-19}$$

$$\bar{S}_{nm} = \frac{1}{K_{nm}} S_{nm} \tag{B.1-20}$$

The reader is cautioned that other definitions of normalization (different from that given above) are used in specific application areas -- for example, the study of the geomagnetic field.

UNIT ONE
REVIEW EXERCISES

Chapter Two

(Section 1.2.1)

1. At what latitude is the difference between geodetic and geocentric latitude a maximum?

(Section 1.2.2)

2. A triangulation survey starts with a baseline (running west to east) of 2000 m connecting points P_0 and P_1. At point P_0, the angle between the baseline and an unknown point, P_2, is measured as 44 deg. At point P_1, the angle between the baseline and point P_2 is measured to be 106 deg. Using P_0 as origin, what are the coordinates of point P_2?

3. Continuing the triangulation survey of Exercise 2, the line P_1P_2 is used as a baseline to survey a new point, P_3. At P_1, the angle between the baseline (P_1P_2)and point P_3 is 52 deg. At P_2, the angle between the baseline and P_3 is 88 deg. Find the coordinates of P_3 relative to the origin, P_0.

4. A trilateration survey starts with a baseline from the origin, P_0, to an initial point, P_1, with length 900 m and azimuth (measured positive from north toward east) of 30 deg. The following distances are measured:

FROM	TO	MEASURED DISTANCE (m)
P_1	P_2	1000
P_0	P_2	1100
P_2	P_3	1200
P_1	P_3	1800

Find the coordinates of P_1, P_2, and P_3 relative to the origin (P_0).

5. An open traverse (refer to Fig. 1.2-9) begins at the origin, P_0, along an initial baseline with azimuth 75 deg and length 1600 m, connecting P_0 with P_1. The traverse continues as shown:

FROM	TO	MEASURED DISTANCE (m)	MEASURED ANGLE (deg)
P_1	P_2	2000	160
P_2	P_3	800	200
P_3	P_4	2500	250

Angles are measured in a clockwise sense from the previous traverse line. Find the coordinates of points P_1, P_2, P_3, and P_4 relative to the origin (P_0).

6. Referring to Exercise 5, suppose that distance measurements are correct to within 0.1 percent, while angles are correct to within 1.0 min. Find upper bounds for the position error of point P_4.

(Section 1.2.3)

7. Describe the factors that limit the accuracy of maps prepared by photogrammetric techniques from overlapping aerial photographs.

(Section 1.2.4)

8. It is stated in the text (Section 1.2.4.1) that "a perfect atomic clock would not, after a lapse of many years, correctly predict such phenomena as sunrise, sunset, star transits, eclipses, etc." Explain why this is the case.

9. Using the Star Catalog section reproduced as Table 1.2-7 (Section 1.2.4.2), calculate the angular distance between the stars α CMa (Sirius) and γ Gem.

<u>Chapter Three</u>

10. If an individual weighs exactly 80 kg at sea level on the equator, how much would this person weight at the North Pole?

11. If the earth were to stop rotating, by how much would gravity at the equator change?

12. Compute the excess gravity acceleration (gravity anomaly) at the surface of the earth, caused by a sphere of depleted uranium, with a mass of 100 kg, buried 10 m below the surface. Would the presence of this object be detectable by the use of ordinary gravimeters?

13. Referring to Appendix B as required, verify Fig. 1.3-4b.

14. Referring to Appendix B as required, plot the Legendre polynomial P_3 (sin ϕ') and verify Fig. 1.3-4a.

(Section 1.3.2)

15. What is the normal gravity at latitude 48 deg N using the International Gravity Formula [Eq. (1.3-5)]?

16. What is the normal gravity at latitude 48 deg N using the Geodetic Reference System 1967 gravity formula?

17. At a point at sea level at latitude 48 deg N, there is a measured gravity of 980933.8 mgal. What is the gravity anomaly at that point?

(Section 1.3.3)

18. Suppose that a gravity measurement is to be made using a pendulum. If the pendulum's period is to be measured very precisely, how accurately must the dimensional stability of the pendulum's length be maintained for measurement accuracy of one mgal? one μgal?

19. In Exercise 18, what is the timing accuracy required to be commensurate with a one mgal level of dimensional stability.

20. Suppose that a pendulum gravimeter has both a one part per million (RMS) dimensional error and an RMS timing error as computed in the previous exercise. If both of these errors are random and are unrelated, what is the RMS error in measured gravity?

UNIT ONE
READING LIST

Chapter One

Geodesy: Trends and Prospects, Committee on Geodesy, Assembly of Mathematical and Physical Sciences (National Research Council), National Academy of Sciences, Washington, D.C., 1978.

Ewing, C.E., and Mitchell, M.M., Introduction to Geodesy, Elsevier, New York, 1970.

(Section 1.2.2)

Mueller, I.I., and Ramsayer, K.H., Introduction to Surveying, Frederick Ungar, New York, 1979.

(Section 1.2.3)

Burnside, C.D., Mapping from Aerial Photographs, Wiley, New York, 1979.

Dickinson, G.C., Maps and Air Photographs, Second Edition, Wiley (Halsted Press), New York, 1979.

Ghosh, S.K., Analytical Photogrammetry, Pergamon Press, New York, 1979.

(Section 1.2.4)

The American Ephemeris and Nautical Almanac, Nautical Almanac Office, United States Naval Observatory, U.S. Government Printing Office, Washington, D.C. (Annual).

Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac, H.M. Nautical Almanac Office, H.M. Stationery Office, London, 1977.

McNally, D., Positional Astronomy, Wiley (Halsted Press), New York, 1977.

Smart, W.M., Textbook on Spherical Astronomy, (Sixth Edition), Cambridge University Press, New York, 1977.

Mueller, I.I. Spherical And Practical Astronomy As Applied to Geodesy, Frederick Ungar, New York, 1969.

(Section 1.2.6)

Richardus, P., and Adler, R.K., Map Projections, North-Holland/ American Elsevier, New York, 1972.

Greenwood, D., Mapping, University of Chicago Press, 1964.

Chapter Three

Heiskanen, W.A., and Moritz, H., Physical Geodesy, Freeman, San Francisco, 1967.

Chapter Four

Kaula, W.M., Theory of Satellite Geodesy, Blaisdell Publishing Company, Waltham, Mass., 1966.

Sterne, T.E., An Introduction To Celestial Mechanics, Interscience, New York, 1960.

Roy, A.E., Orbital Motion, Wiley (Halsted Press), New York, 1978.

Herrick, S., Astrodynamics, Van Nostrand/Reinhold, New York, Vol. 1, 1971, Vol. 2, 1972

Appendix A

Kellogg, O.D., Foundations Of Potential Theory, Dover, New York, 1953 (Reprint).